Numer. Math. (2010) 115:45–69
DOI 10.1007/s00211-009-0270-2

Numerische Mathematik

An algorithm for computing solutions of variational problems with global convexity constraints

Ivar Ekeland · Santiago Moreno-Bromberg

Received: 12 April 2008 / Revised: 9 June 2009 / Published online: 15 October 2009

Abstract We present an algorithm to approximate the solutions to variational problems where set of admissible functions consists of convex functions. The main motivation behind the numerical method is to compute solutions to Adverse Selection problems within a Principal-Agent framework. Problems such as product lines design, optimal taxation, structured derivatives design, etc. can be studied through the scope of these models. We develop a method to estimate their optimal pricing schedules.

Mathematics Subject Classification (2000) 49-04 · 49M25 · 49M37 · 65K10 · 91B30 · 91B32

1 Introduction

Isaac Newton was the first person to state and solve a variational problem with convexity constraints. In the *Principia*, he sought the shape of convex solid that encounters the least resistance when moving through a fluid. The problem can be stated as follows. Let Θ be a smooth, convex subset of $\mathbb{R}^2$. Define

We thank Guillaume Carlier and Yves Lucet for their thoughtful comments and suggestions. We would also like to acknowledge two anonymous referees and the editor, whose thorough reading and suggestions led to an improved version of our original manuscript.

I. Ekeland
Department of Mathematics, UBC, Canada Research Chair in Mathematical Economics,
1984 Mathematics Road, Vancouver, BC V6T 1Z2, Canada
e-mail: ekeland@math.ubc.ca

S. Moreno-Bromberg (✉)
Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6,
10099 Berlin, Germany
e-mail: smoreno@math.hu-berlin.de

$$I[v] := \int\limits_{\Theta} \frac{d\theta}{1 + |\nabla v|^2}, \quad \text{and} \quad \mathcal{C} := \{v : \Theta \to \mathbb{R} \mid v \text{ is convex}\}.$$

One seeks the v in $\mathcal{C}$ which minimize(s) I. It should be noted that even if the Lagrangian satisfied the necessary coercivity properties (see for example [11, Sect. 8]), the restriction $v \in \mathcal{C}$ would make it quite difficult to use the Euler–Lagrange equations, which are satisfied only when the constraints are not binding. Newton solved the problem by assuming, quite naturally, that the function (and the domain) were radially symmetric. Four centuries later, Brock et al. [4] (see also [15]) showed that, if one removes the symmetry assumption on the solution, one gets a lower minimum, and this result sparked new interest to the study of variational problems with convexity constraints. At the same time, such problems were also cropping up in finance and economics, because of concerns with asymmetry of information, particularly adverse selection (see [3] for a comprehensive account). Typically, when a monopolist addresses a market consisting of consumers with different tastes and means, he/she will not sell a single product, but will devise a line of products with different qualities and prices, each of which addresses a segment of the market. These products then compete with each other (even if I am well off, I may want to buy the product devised for less wealthy people, even though the quality is less, because I find it a better bargain). The monopolist's problem (also known in the economic literature as the principal-agent problem, the monopolist being the principal and the consumers the agents) is to devise a pricing schedule such that his/her profit is maximal. This is called non-linear pricing. The field took off in 1978 with a seminal paper of Mussa and Rosen [17], and since then has produced a considerable stream of contributions ([2,7,19], etc.).

In models where goods are described by a single quality and the set of consumers is differentiated by a single parameter, it is in general possible to find closed form solutions for the pricing schedule, which is precisely the mathematical content of the paper by Mussa and Rosen. This is, however, not the case when multidimensional qualities and consumer types are considered. The question was first addressed by Rochet and Choné [19], who provided conditions for the existence of an optimal pricing rule and fully characterized the ways in which the monopolist discriminates among consumers in a multidimensional setting. They pointed out that it is only in very special cases that one can expect to find closed form solutions. The same holds true for models where the set of goods lies in an infinite-dimensional space, even when agent types are one-dimensional. This framework was first used, to our knowledge, by Carlier et al. [7] to price OTC (over-the-counter) financial derivatives. It was then extended by Horst and Moreno-Bromberg [12] to model the actions of a monopolist who has an initial risky position that he/she evaluates via a coherent risk measure, and who intends to transfer part of his/her risk to a set of heterogenous agents. In both cases the authors find that only very restrictive examples allow for explicit solutions.

Given that a great variety of problems, such as product lines design, optimal taxation, structured derivatives design, translate into variational problems with global convexity constraints, there is a clear need for robust and efficient numerical methods that approximate their solutions. Note here a particular requirement, which is an additional burden on the numerics: the quantity of interest is not the solution itself, i.e. the maximizing function, but its gradient. Indeed, if $f(x)$ is the solution, then $\nabla f(x)$ is the quality bought by consumers of type x.

The main difficulty is to find an approximation of the cone of convex functions. The first attempt in this direction, due to Kawohl and Schwab [13], consisted of using a first-order finite elements method. Unfortunately it turned out to be flawed: indeed, Choné and Le Meur [9] proved that, given a family of structured meshes M_h, one can always find a convex function u that is not a limit of convex functions u_h, with u_h piecewise linear on M_h. In other words, internal approximations with first-order finite elements are bound to fail, and the authors illustrate their point by numerical examples. Carlier et al. [8] then introduced an external approximation by first-order finite elements, and showed that their algorithm converges when the functional to be maximized (or minimized) is quadratic and there are no constraints. This is not the case for the Rochet–Choné problem nor for most problems arising from economic theory, to which this method does not provide an answer. The editor has drawn our attention to a very interesting paper by Anguilera and Morin [1], which appeared while this one was being revised, and which takes yet another approach. Anguilera and Morin consider first- and second-order finite differences on a given family of structured meshes M_h, which enables them do define a discrete gradient and a discrete Hessian; they then consider the class of functions such that the discrete Hessian is non-negative, and prove that any convex function can be approximated in this way. On each mesh, the discretized problem is solved using semi-definite programming (SDP), and when h goes to 0 the corresponding solutions converge to the theoretical solution.

Our approach differs from the preceding ones. It is an interior method: at each step the approximate minimizer is convex. This is important from the economic point of view because, as we mentioned earlier, the convexity condition embodies the information constraint which any proposed contract must satisfy: non-convex functions represent contracts which cannot be implemented (i.e. agents will not behave in the way the principal expects). However, it is not a finite element method, nor does it use finite differences or SPD: it relies on the well-known fact that a convex function is the supremum of its affine minorants:

$$u(x) = \sup_{(a,x^*)} \left\{ \langle x, x^* \rangle + a | \forall y,\ \langle y.x^* \rangle + a \leq u(y) \right\}$$

or, equivalently, of all its tangent hyperplanes:

$$u(x) = \sup_{y} \left\{ \langle x - y, u'(y) \rangle + u(y) \right\}$$

We will replace the preceding formula by the following:

$$u_h(x) = \sup \left\{ \langle x - y, u'(y) \rangle + u(y) | y \in M_h \right\}$$

where M_h is a given mesh. It is clear that the function u_h is convex, and we will show that, under reasonable assumptions, it converges to u as the step h goes to 0. The function u_h is piecewise linear, in the sense that it is linear on certain cells downstairs, but these cells are unrelated to the mesh M_h we started from; in particular, they do not have points in the mesh as vertices. In fact, the shape of these cells can be quite complicated, as illustrated in Figs. 5 and 6 of Sect. 4 (although, of course, they are convex polyhedra). In this sense our method differs from a finite-elements method, where the given mesh determines the cells on which the approximate solution is linear (or rather, affine). This is why it evades the criticism of Choné and Le Meur: since the decomposition into cells is not given a priori, but adapts to the solution, there cannot be the same geometrical obstruction that befalls a decomposition where the shapes are defined a priori. The same reason may also explain why there are no geometrical obstruction either arising from the fact that our method uses simultaneously the function u and its gradient ∇u: since the decomposition into cells is unstructured, and adapts to the solution, it will also adapt to fit any integrability conditions. In Lachand-Robert and Oudet [14], have used a similar idea in a more geometrical setting: they seek a convex body that minimizes a certain functional, and they proceed by starting from an admissible polyhedron and iteratively modifying the normals to the facets in order to find an approximate minimizer.

We first apply our method to a benchmark problem, taken from Choné and Le Meur [9], where the solution can be computed explicitly, and we find that our method works in situations where they find they do not have convergence. We then estimate the minimizers for the well known Rochet–Choné problem (see [19]), where we find that our solution matches the theoretical one, to a problem in OTC pricing of securities [7], and to a risk-minimization problem as in [12]. Our algorithm proves to be versatile enough to provide approximate solutions even when the criterion is non-convex, as in [7]. This is of particular interest for economic situations, since convexity goes hand-in-hand with the assumption that the agents have quasi-linear preferences, which in many cases is simply too restrictive.

The remainder of this paper is organized as follows. In Sect. 2, we state our problem and provide our notation. Our algorithm and a proof of its convergence are presented in Sect. 3, together with further discussion of the method. In Sect. 4, we show the solutions obtained via our algorithm to several problems found in the literature.[1] Since these problems share a common microeconomic motivation, we include a brief discussion on the latter. The examples include the well known Rochet–Choné problem, a one dimensional example from Carlier, Ekeland and Touzi and the risk transfer case of Horst and Moreno-Bromberg for a principal who offers call options with type-dependent strikes and evaluates his/her risk via the "shortfall" of his/her position. This section is followed by our conclusions. Finally, the appendix is devoted to technical results.

2 Setting

Throughout this paper, we use the following notations:

- $\Theta, Q \subset \mathbb{R}^n$ are convex and compact sets,
- μ denotes the Lebesgue Measure on $\mathbb{R}^n$,
- f is a (probability) density on Θ,
- $L(\theta, z, p) = z - \theta \cdot p + C(p)$, where C is strictly convex and $\mathcal{C}^1$,

[1] Our Matlab codes can be downloaded from: http://www2.hu-berlin.de/math-finance/?q=node/72.

- $\mathcal{C} := \{v : \Theta \to \mathbb{R} \mid v \geq 0$ is convex, and $\nabla v \in Q$ a.e$\}$,
- $I[v] := \int_{\Theta} L(\theta, v(\theta), \nabla v(\theta)) f(\theta) d\theta$.

Our objective is to (numerically) estimate the solution to

$$\mathcal{P} := \inf_{v \in \mathcal{C}} I[v]. \tag{1}$$

Having in mind examples as the one presented in Sect. 4.1, we will refer to C as the *cost function*. See Sect. 4.1 for the microeconomic motivation to the particular structure of the Lagrangian. Given the properties of L and $\mathcal{C}$ we immediately have the following

Proposition 2.1 *Assume $\overline{v}$ solves $\mathcal{P}$, then there is θ_0 in Θ such that $\overline{v}(\theta_0) = 0$.*

Proof Let $\overline{v}_0 = \min_{\theta \in \Theta} \overline{v}(\theta)$ (recall Θ is compact)and define $\overline{u}(\theta) := \overline{v}(\theta) - \overline{v}_0$, then

$$I[\overline{u}] = \int_{\Theta} (\overline{u}(\theta) - \theta \cdot \nabla \overline{u}(\theta) + C(\nabla \overline{u}(\theta))) f(\theta) d\theta = I[\overline{v}] - \overline{v}_0 < I[\overline{v}].$$

This would contradict the hypothesis of $\overline{v}$ being a minimizer of I over $\mathcal{C}$ unless $\overline{v}_0 = 0$. □

Remark 2.2 Notice that Proposition 2.1 would hold for any Lagrangian of the form $L(\theta, z, p) = G(\theta, p) - \theta \cdot z$ for any strictly convex function G that does not depend on z.

It follows from Proposition 2.1 that we can redefine $\mathcal{C}$ to include only functions that have a root in Θ. This, together with the compactness of Q, implies the following proposition, which we will use frequently.

Proposition 2.3 *There exists $0 < K < \infty$ such that $v \leq K$ for all v in $\mathcal{C}$.*

It follows from Proposition 2.3 and the restriction on the gradients that for each choice of cost function C, problem $\mathcal{P}$ has a unique solution, since the functional I will be strictly convex, lower semicontinuous and the admissible set is closed and bounded (see [10]).

Remark 2.4 Our algorithm will still work with more general L's as long as one can prove that the family of feasible minimizers is uniformly bounded.

3 The algorithm

This section contains a detailed description of our algorithm (Sect. 3.1), as well as a proof of its convergence (Sect. 3.2). From this point on, whenever we use superscripts we refer to vectors. For example $V^k = (V_1^k, \ldots, V_m^k)$. On the other hand a subscript indicates a function to be evaluated over some closed, convex subset of $\mathbb{R}^n$ of non-empty interior, ie, $\{V_j\}$ is a sequence of functions $V_j : X \to \mathbb{R}$ for some X contained in $\mathbb{R}^n$. We will only consider the domain $\Theta = [a, b]^n$. This choice is made for computational simplicity and it plays no role in our proof of convergence.

3.1 Description

In order to find a interior approximation to the solution to $\mathcal{P}$ we proceed as follows:

1. We partition Θ into Σ_k, which consists of k^n equal cubes of volume $\mu(\Sigma_k) := (\frac{b-a}{k})^n$. The elements of the partition Σ_k will be denoted by σ_j^k, $1 \le j \le k^n$. Then we define Θ_k as the set of centers of the σ_j^k's. The elements of Θ_k will be denoted by θ_j^k. The choice of a uniform partition is done for computational simplicity.
2. We denote $f_i^k = \int_{\sigma_i^k} f(\theta)d\theta$ and associate such weight with θ_i^k.
3. We associate to each element θ_i^k of Θ_k a non-negative number v_i^k and an n-dimensional vector D_i^k. The former represents the value of $v(\theta_i^k)$ and the latter $\nabla v(\theta_i^k)$.
4. We solve the (non-linear) program

$$\mathcal{P}_k := \inf \mu(\Sigma_k) \sum_{i=1}^{k^n} L(\theta_i^k, v_i^k, D_i^k) f_i^k \tag{2}$$

   over the set of all vectors of the form $v = (v_1, \dots, v_{k^n})$ and all matrices of the form $D = (D_1, \dots, D_{k^n})$ such that:
   (a) $v \ge 0$ (non-negativity),
   (b) $D_i \in Q$ for $i = 1, \dots k^n$ (feasibility),
   (c) $v_i - v_j + D_i \cdot (\theta_j - \theta_i) \le 0$ for $i, j \in \{1, \dots, k^n\}$ $i \ne j$ (convexity).
   If the problem in hand includes Dirichlet boundary conditions these can be included here as linear constraints that the D_i^k's corresponding to points on the "boundary" of Θ_k must satisfy.
5. Let $(\overline{v}^k, \overline{D}^k)$ be the solution to $\mathcal{P}_k$. Define $\overline{v}_k(\theta) := \max_i p_i(\theta)$, where

$$p_i(\theta) = \overline{v}_i^k + \overline{D}_i^k \cdot (\theta - \theta_i^k).$$

6. $\overline{v}_k$ yields an approximation to the minimizer of $\mathcal{P}$.

Remark 3.1 The constraints of the non-linear program determine a convex set. Notice that the number of constraints associated to problem $\mathcal{P}_k$ over $[a, b]^n$ is $k^n + nk^n + k^n(k^n - 1)$. The summands correspond to the positivity, feasibility and convexity constraints respectively. Hence, this number grows polynomially with the number of elements in the lattice and exponentially with dimension.

Remark 3.2 4 (c) guarantees that p_i is a supporting hyperplane of the convex hull of the points $\{(\theta_1, v_1), \dots, (\theta_{k^n}, v_{k^n})\}$. Note that $\overline{v}_k$ is a piecewise affine convex function, but that the cells on which it is affine are unrelated to the cells of partition Σ_k. We are not using a finite-elements method. This will be discussed further in Sect. 3, where examples will be given.

It should be noted that the evaluation of $\overline{v}_k$ is non-local. This has the drawback of being numerically very expensive; however, it yields intrinsically convex functions. Moreover, since these convex functions are lower semicontinuous (they are the pointwise supremum of linear functions) and finite, the algorithm produces continuous functions.

3.2 Convergence

In this section, we prove convergence of the algorithm. We denote, for notational convenience

$$J_k(v^k, D^k) := \sum_{i=1}^{k^n} \left(v_i^k - \theta_i \cdot D_i^k + C(D_i^k) \right) \mu(\Sigma_k) f_i^k.$$

Proposition 3.3 *Under the assumptions made on L, the problem $\mathcal{P}_k$ has a unique solution.*

Proof The mapping

$$(v^k, D^k) \to J_k(v^k, D^k)$$

is strictly convex. It follows from Proposition 2.3 that any feasible vector-matrix pair (v^k, D^k) must lie in $[0, K]^k \times Q^k$, which together with Remark 3.1 implies $\mathcal{P}_k$ consists of minimizing a strictly convex function over a compact and convex set. The result then follows from general theory (see, for example [10]). □

Proposition 3.4 *There exists $\overline{v} \in \mathcal{C}$ such that*:

1. *The sequence $\{\overline{v}_k\}$ generated by the $\mathcal{P}_k$'s has a subsequence $\{\overline{v}_{k_j}\}$that converges uniformly to $\overline{v}$.*
2. $\lim_{k_j \to \infty} I[\overline{v}_{k_j}] = I[\overline{v}]$.

Proof The bounded (Proposition 2.3) family $\{\overline{v}_j\}$ is uniformly equicontinuous, as it consists of convex functions with uniformly bounded subgradients (they are required to lie in Q). By the Arzela–Ascoli theorem we have that, passing to a subsequence if necessary, there is a non-negative and convex function $\overline{v}$ such that

$$\overline{v}_k \to \overline{v} \quad \text{uniformly on } \Theta.$$

By convexity $\nabla \overline{v}_k \to \nabla \overline{v}$ almost everywhere (Lemma A.4); since $\nabla \overline{v}_k(\theta)$ belongs to the bounded set Q, the integrands are dominated. Therefore, by Lebesgue Dominated Convergence we have

$$\lim_{k \to \infty} I\,[\overline{v}_k] = I[\overline{v}].$$

□

Let $\overline{u}$ be the maximizer of $I[\cdot]$ within $\mathcal{C}$. Our aim is to show that $\{\overline{v}_k\}$ is a minimizing sequence of problem $\mathcal{P}$, in other words that

$$\lim_{k \to \infty} I\,[\overline{v}_k] = I\,[\overline{u}]\,.$$

To this end, we need the following auxiliary definition:

Definition 3.5 Let $\overline{u}$ be such that $\inf_{u\in\mathcal{C}} I[u] = I[\overline{u}]$. Given Θ_k, define for $i = 1,\ldots,k^n$:

1. $\overline{u}_i^k := \overline{u}(\theta_i^k)$,
2. $G_i^k := \nabla u(\theta_i^k)$,
3. $q_i(\theta) := \overline{u}_i^k + G_i^k \cdot (\theta - \theta_i^k)$, and
4. $\overline{u}_k(\theta) := \max_i q_i(\theta)$.

Notice that $\overline{u}_k(\theta)$ is also constructed as the convex envelope of a family of affine functions. The inequalities

$$J_k(\overline{u}^k, G^k) \geq J_k(\overline{v}^k, \overline{D}^k), \tag{3}$$

and

$$I[\overline{v}_k] \geq I[\overline{u}] \tag{4}$$

follow from the definitions of $J_k(\overline{v}^k, \overline{D}^k)$, $\overline{v}_k$ and $\overline{u}_k$, as does the following

Proposition 3.6 *Let $\overline{u}$ and $\overline{u}_k$ be as above, then $\overline{u}_k \to \overline{u}$ uniformly as $k \to \infty$.*

Proposition 3.7 *For each k there exist $\epsilon_1(k)$ and $\epsilon_2(k)$ such that*

$$\left| J_k(\overline{v}^k, \overline{D}^k) - I[\overline{v}_k] \right| \leq \epsilon_1(k), \tag{5}$$

$$\left| J_k(\overline{u}^k, G^k) - I[\overline{u}_k] \right| \leq \epsilon_2(k), \tag{6}$$

and $\epsilon_1(k), \epsilon_2(k) \to 0$ as $k \to \infty$.

Proof We will show inequality (5) holds, the proof for (6) is analogous. Define the simple function

$$w_k(\theta) := L(\theta_j^k, \overline{v}_j^k, \overline{D}_j^k),\ \ \theta \in \sigma_j^k,$$

hence

$$J_k(\overline{v}^k, \overline{D}^k) = \int_\Theta w_k(\theta)d\theta. \tag{7}$$

The left-hand side of (5) can be written as

$$\left| \int_\Theta w_k(\theta)d\theta - I[\overline{v}_k] \right|. \tag{8}$$

It follows from Lemma A.6 that there exists $\epsilon_1(k)$, such that

$$\left| \int_{\Theta} w_k(\theta) d\theta - I[v_k] \right| \leq \epsilon_1(k),$$

and

$$\epsilon_1(k) \to 0 \quad \text{when} \quad k \to \infty.$$

□

We can now prove our main theorem, namely

Theorem 3.8 *The sequence* $\{\overline{v}_k\}$ *is minimizing for problem* $\mathcal{P}$.

Proof It follows from Proposition 3.7 and Eq. (3) that

$$I[\overline{u}_k] + \epsilon_2(k) + \epsilon_1(k) \geq I[\overline{v}_k] \geq I[\overline{u}]. \tag{9}$$

Letting $k \to \infty$ in (9) and using Proposition 3.4 yields the desired result. □

Remark 3.9 In Choné and Le Meur [9] (see also Carlier et al. [8]), it was pointed out that, when using finite elements on regular meshes, one encounters a geometrical obstruction. Namely, let T_n be a sequence of quasiuniform regular triangulations T_n, and suppose there are two vectors h and k such that, for every unit vector ν which is normal to an edge of each element in the triangulation

$$\langle \nu, h \rangle \cdot \langle \nu, k \rangle \geq 0.$$

Then, if u is the limit in L^∞_{loc} of a sequence of convex functions u_n which are piecewise linear on the triangulation, we must have:

$$\frac{\partial^2 u}{\partial h \partial k} \geq 0.$$

As Figs. 2 and 3 in the next section will show, the partition generated by our method is not a triangulation, nor is it regular, so it is not affected by this geometrical obstruction.

4 Examples

In this section, we show some results of implementing our algorithm. The first three examples reduce quadratic programs, whereas the fourth and fifth ones are non-linear optimization programs. All the computer coding has been written in Matlab. However, the supplemental Tomlab 6.0 Optimization Toolbox was used to speed up running times. To develop Examples 4.2, 4.3 and 4.4 we have used the drop-in replacement for *quadprog* in Matlab's Optimization toolbox. We have also used the drop-in replacement for *fmincon* in Example 4.5. To solve the non-linear program in Example 4.6 we used Tomlab's solver *conSolve*. In all of our examples the *Exit Flag* returned was 0. In other words, the iteration points were close. These examples, which have been taken from [7,9,12,19], share a common microeconomic motivation, for which we provide an overview. We refer the interested reader to [3] for a comprehensive presentation of Principal-Agent models and Adverse Selection, as well as multiple references.

4.1 Some microeconomic motivation

Consider an economy with a single PRINCIPAL and a continuum of AGENTS. The latter's preferences are characterized by n-dimensional vectors. These are called the agents' TYPES. The set of all types will be denoted by $\Theta \subset \mathbb{R}^n$. The individual types θ are private information, but the principal knows their statistical distribution, which has a density $f(\theta)$.

We assume goods are characterized by (n-dimensional) vectors describing their utility-bearing attributes. The set of TECHNOLOGICALLY FEASIBLE goods that the principal can deliver will be denoted by $Q \subset \mathbb{R}^n_+$, and it will be assumed to be compact and convex. The cost to the principal of producing one unit of product q is denoted by $C(q)$. Products are offered on a take-it-or-leave-it basis, each agent can buy one or zero units of a single product q and it is assumed there is no second-hand market. The (type-dependent) preferences of the agents are represented by the function

$$U : \Theta \times Q \to \mathbb{R}.$$

The (non-linear) price schedule for the technologically feasible goods is represented by

$$\pi : Q \to \mathbb{R}.$$

When purchasing good q at a price $\pi(q)$ an agent of type θ has net utility

$$U(\theta, q) - \pi(q).$$

Each agent solves the problem

$$\max_{q \in Q} \{U(\theta, q) - \pi(q)\}.$$

By analyzing the choice of each agent type under a given price schedule π, the principal (partially) screens the market. Let

$$v(\theta) := U(\theta, q(\theta)) - \pi(q(\theta)), \tag{10}$$

where $q(\theta)$ belongs to $argmax_{q\in Q}\{U(\theta,q)-\pi(q)\}$. Notice that for all q in Q we have

$$v(\theta)\geq U(\theta,q)-\pi(q). \tag{11}$$

Analogous to the concepts of SUBDIFFERENTIAL and CONVEX CONJUGATE from classical Convex Analysis, we have that the subset of Q where (11) is an equality is called the U-SUBDIFFERENTIAL of v at θ and v is the U-CONJUGATE of π (see, for example, [6]). We write

$$v(\theta)=\pi^U(\theta),$$

and

$$\partial_U v(\theta):=\{q\in Q\ \mid\ \pi^U(\theta)+\pi(q)=U(\theta,q)\}.$$

To simplify notation let $\pi(q(\theta))=\pi(\theta)$. A single pair $(q(\theta),\pi(\theta))$ is called a CONTRACT, whereas $\{(q(\theta),\pi(\theta))\}_{\theta\in\Theta}$ is called a CATALOGUE. A catalogue is called INDIVIDUALLY RATIONAL if $v(\theta)\geq v_0(\theta)$ for all $\theta\in\Theta$, where $v_0(\theta)$ is type's θ non-participation (or reservation) utility. We normalize the reservation utility of all agents to zero, and assume there is always an OUTSIDE OPTION q_0 that denotes non-participation. Therefore we will only consider functions $v\geq 0$. The Principal's aim is to devise a pricing function $\pi:Q\rightarrow\mathbb{R}$ as to maximize his/her income

$$\int_\Theta (\pi(\theta)-C(q(\theta)))\,f(\theta)d\theta. \tag{12}$$

Inserting (10) into (12) we get the alternate representation

$$\int_\Theta (U(\theta,q(\theta))-v(\theta)-C(q(\theta)))\,f(\theta)d\theta. \tag{13}$$

Expression (13) is to be maximized over all pairs (v,q) such that v is U-convex and non-negative and $q(\theta)\in\partial_U v(\theta)$. Characterizing $\partial_U v(\theta)$ in a way that makes the problem tractable can be quite challenging. In the case where $U(\theta,q(\theta))=\theta\cdot q(\theta)$, as in [19], for a given price schedule $\pi:Q\rightarrow\mathbb{R}$, the indirect utility of an agent of type θ is

$$v(\theta):=\max_{q\in Q}\{\theta\cdot q-\pi(q)\}. \tag{14}$$

Since v is defined as the supremum of its affine minorants, it is a convex function of the types. It follows from the Envelope Theorem that the maximum in Eq. (14) is attained if $q(\theta)=\nabla v(\theta)$, and we may write

$$v(\theta)=\theta\cdot\nabla v(\theta)-\pi(\nabla v(\theta)). \tag{15}$$

The principal's aggregate surplus is given by

$$\int_{\Theta} (\pi(q(\theta)) - C(q(\theta)))\, f(\theta) d\theta. \tag{16}$$

Inserting (15) into (16) we get that the principal's objective is to maximize

$$I[v] := \int_{\Theta} (\theta \cdot \nabla v(\theta) - C(\nabla v(\theta)) - v(\theta))\, f(\theta) d\theta \tag{17}$$

over the set

$$\mathcal{C} := \{v : \Theta \to \mathbb{R} \mid v \,\text{convex}, \ v \geq 0, \ \nabla v(\theta) \in Q\}.$$

4.2 A benchmark

Following in the footsteps of Choné and Le Meur [9], we use the following example as a benchmark. Within this section, we use the notation

$$\mathbf{C} := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

where $\rho \in (-1, 1)$. Our aim is to study the problem of minimizing

$$I[v] = \int_{[0,1]^2} \left(\frac{1}{2} \nabla v(\theta)^t \mathbf{C}\, \nabla v(\theta) - \theta \cdot \nabla v(\theta) \right) d\theta$$

over the set of convex functions with positive partial derivatives. The main interest behind approximating the solution to this problem is that the "true" solution can be given explicitly. To do so, one first shows that the Euler–Lagrange equations for the above problem have, up to an additive constant, the solution

$$\tilde{v}(\theta_1, \theta_2) = \frac{1}{2(1-\rho^2)} \left(\theta_1^2 - 2\rho\theta_1\theta_2 + \theta_2^2 \right). \tag{18}$$

It follows from general theory (see for example [11, Chapter 8]) that for $\mathcal{A} = \mathbb{H}^2([0, 1]^2)$, the problem $\inf_{\mathcal{A}} I[v]$ has a unique solution. Next, one observes that for $|\rho| < 1$, $\tilde{v}$ satisfies the convexity constraint and the positivity constraints on the partial derivatives on $[0, 1]^2$. Therefore Eq. (18) is the solution to the convexly-constrained problem. We proceed to compare the performance of our algorithm against $\tilde{v}$. A description of how the quadratic program is setup can be found in Sect. 4.3, the

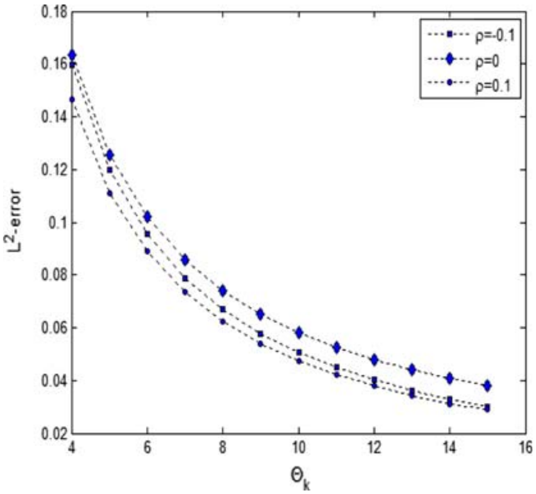


Fig. 1 The L^2-error for $\rho = -0.1, 0$ and 0.1

only difference being that for partition Θ_k, the the four $k^2 \times k^2$ blocks towards the Southeast corner of matrix H are of the form

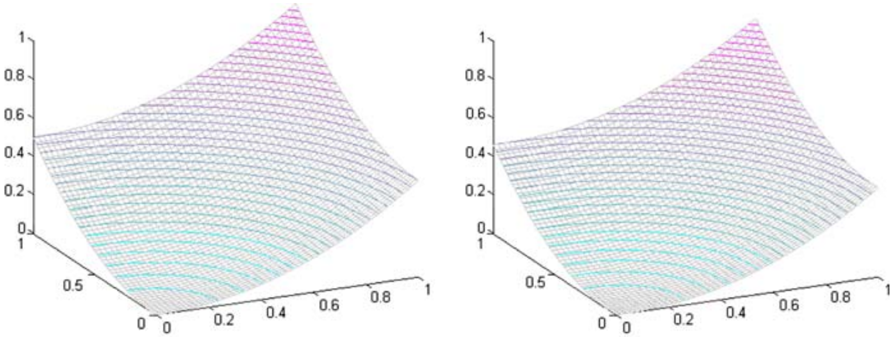
$$\begin{pmatrix} \mathbb{I} & \rho \cdot \mathbb{I} \\ \rho \cdot \mathbb{I} & \mathbb{I} \end{pmatrix},$$

where $\mathbb{I}$ stands for the identity matrix of size $k^2 \times k^2$. For the error estimates we have used the L^2 norm with a uniform discretization of the domain consisting of 10^6 points. In contrast to what was found by Choné and Le Meur in [9], our method exhibits good convergence behaviour for the case $\rho < 0$. In the case $\rho \geq 0$, both out method and the examples in their paper show an adequate reduction of the error as the partitions (or meshes) become finer. The running times for the Θ_{15} case and $\rho = -0.1, 0$ and 0.1 were: 452.12 s, 388.21 s and 437.74 s, respectively (Fig. 1). These were obtained running Matlab Release 2007a on a computer with an AMD Athlon 64 × 2 Dual-Core Processor running at 1.80 GHz and with 2 MB of RAM. It is interesting to note that about 30 s are devoted to the construction of the affine envelopes, while the rest goes into solving the quadratic programs. Finally we present, for the case $\rho = 0$, the plot of $graph\{\tilde{v}\}$ in Fig. 2a and the plot of $graph\{\bar{v}\}$ (the output of our algorithm) in Fig. 2b. We used a 14×14 lattice to generate Fig. 2b.

4.3 The Rochet–Choné problem

In this example, we test our algorithm on the Roche–Choné problem (also known as the Mussa–Rosen problem in a square). A description of the solution to this problem was found by Rochet and Choné in [19], and the output of our algorithm matches it in a satisfactory fashion. The following structures are shared with Example 4.4:

- $x = (v^k, D^k)$, this structure will determine any possible candidate for a minimizer to $J_k(\cdot, \cdot)$ in the following way: v^k is a vector of length k^2 that will contain the



(a) The graph of $\tilde{v}$. **(b)** The graph of $\bar{v}$.

Fig. 2 The true versus the approximate solutions for $\rho = 0$

approximate values of the optimal function $\overline{v}$ evaluated on the points of the lattice. The vector D^k has length $2 * k^2$ and it contains what will be the partial derivatives of $\overline{v}$ at the same points.

- h is a vector of length $3 * k^2$. The product $h \cdot x$ provides the discrete representation of the integral $\int_\Theta (\theta \cdot \nabla v - v(\theta)) f(\theta) d\theta$.
- B is the matrix of constraints. The inequality $Bx \leq 0$ imposes the non-negativity of v and D and the convexity of the resulting $\overline{v}_k$.

Let $\Theta = [1, 2]^2$, $C(q) = \frac{1}{2}\|q\|^2$, and assume the types are uniformly distributed. In this case we have to solve the quadratic program

$$\sup_x \left\{ h \cdot x - \frac{1}{2} x^t H x \right\}$$

subject to

$$Bx \leq 0.$$

Here H is a $(3 * k^2) \times (3 * k^2)$ matrix whose first k^2 columns are zero, since v does not enter the cost function; the four $k^2 \times k^2$ blocks towards the Southeast corner form a $(2 * k^2) \times (2 * k^2)$ identity matrix. Therefore $\frac{1}{2} x^t H x$ is a the discrete representation of $\int_\Theta \frac{1}{2} \|\nabla v(\theta)\|^2 d\theta$. Figure 3 was produced using a 17×17-points lattice and a uniform density.

In Fig. 4, we compare the region of traded qualities found in [19] (Fig. 4a) to the output of the previous execution of our algorithm (Fig. 4b). Notice that the above execution does a good job of capturing the region of "high quality" products, i.e. those destined to the upper echelon of the market. However, at this level of precision the Southwest corner of the set of fully differentiated products is slightly widened. This is not too surprising and it showcases the difficulty of the problem in hand: one requires not only $\bar{v}$, but also $\nabla \bar{v}$.

Our algorithm generates at each step a new partition of the domain into cells on which the approximate minimizer is affine. Figures 5 and 6 show the cells correspond-

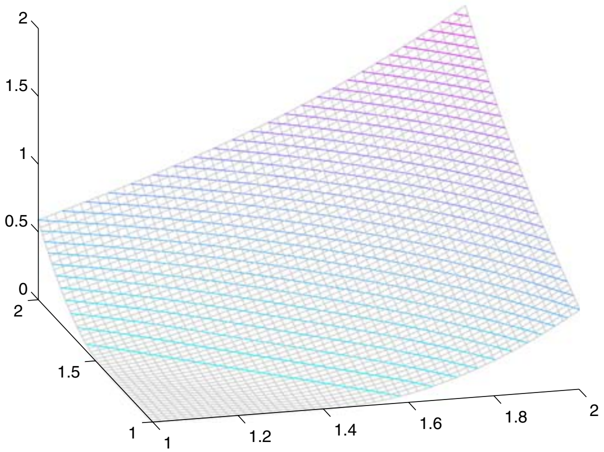


Fig. 3 Solution for uniformly distributed agent types

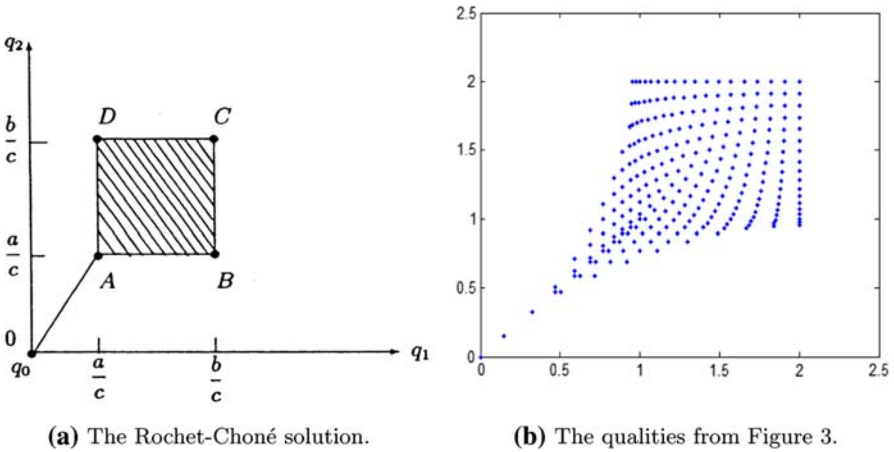


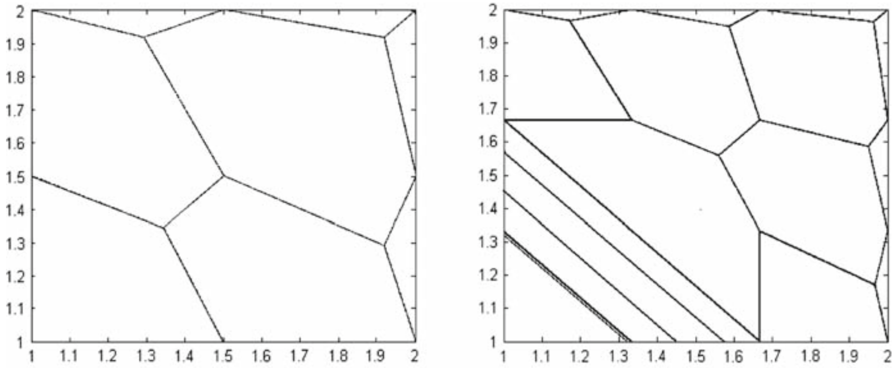
Fig. 4 The traded qualities

ing to the Θ_3 to Θ_6 cases (3×3 to 6×6 partitions of the domain). We stress again that these sub-divisions arise *ex-post*: their characteristics are not known a priori, but follow from the execution of the optimization routine.

4.4 The Mussa–Rosen problem with a non-uniform density

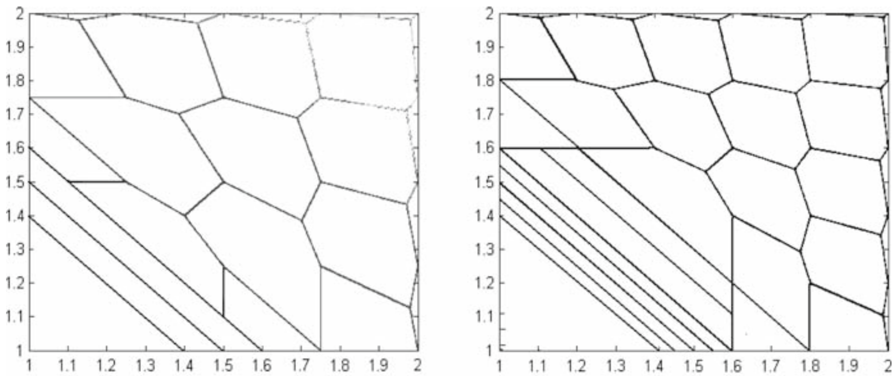
We keep the cost function and the partition of the previous example, but now assume the types are distributed according to a bivariate normal distribution with mean (1.9, 1) and variance-covariance matrix

$$\begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix}$$



(a) Subdivision of the domain for Θ_3. **(b)** Subdivision of the domain for Θ_4.

Fig. 5 Subdivisions of the domain for the 3×3 and 4×4 lattices



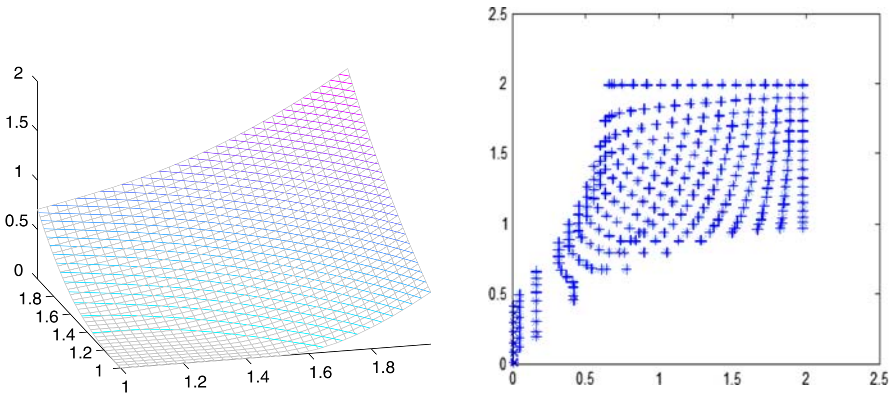
(a) Subdivision of the domain for Θ_5. **(b)** Subdivision of the domain for Θ_6.

Fig. 6 Subdivisions of the domain for the 5×5 and 6×6 lattices

The weight assigned to each agent type is built into h and H, so the vector x remains unchanged. We obtain Fig. 7.

Remark 4.1 It is interesting to see that in this case bunching of the second kind, as described by Roché and Choné in [19], appears to be eliminated as a consequence of the skewed distribution of the agents. This can be seen in the non-linear level curves of the optimizing function v, and it is also quite evident in the plot of the traded qualities.

The codes for the two previous examples were run on Matlab 7.0.1.24704 (R14) in a Sun Fire V480 (4×1.2 HGz Ultra III, 16 GB RAM) computer running Solaris 2.10 OS. In the first example 57.70 s of processing time were required. The running time in the second example was 81.72 s.



(a) The approximating function $\overline{v}_{17}$. **(b)** The qualities traded.

Fig. 7 Optimal solution for normally distributed agent types

4.5 An example with non-quadratic cost

In this example we approximate a solution to the problem of a principal who is selling over-the-counter financial derivatives to a set of heterogeneous agents. This model is presented by Carlier et al. [7]. They start with a standard probability space $(\Omega, \mathcal{F}, P)$, and the types of the agents are given by their risk aversion coefficients under the assumption of mean-variance utilities; namely, the set of agent types is $\Theta = [0, 1]$, and the utility of an agent of type θ when facing product X is

$$U(\theta, X) = \mathrm{E}[X] - \theta \mathrm{Var}[X].$$

Under the assumptions of a zero risk-free rate and access to a complete market by the principal, his/her cost of delivering product $X(\theta)$ is given by $\sqrt{-\xi v'(\theta)}$; where ξ is the variance of the Radon–Nikodym derivative of the (unique) martingale measure, and $\mathrm{Var}[X(\theta)] = -v'(\theta)$. The principal's problem can be written as

$$\sup_{v \in \mathcal{C}} \int\limits_{\Theta} \left(\theta v'(\theta) + \sqrt{-v'(\theta)} - v(\theta) \right) d\theta, \tag{19}$$

where $\mathcal{C} := \{v : \Theta \to \mathbb{R} \mid v \text{ convex}, \ v \geq 0, v' \leq 0 \text{ and Var } [X(\theta)] = -v'(\theta)\}$. Figure 8 shows an approximation of the maximizing $\overline{v}$ using 25 agent types.

4.6 Minimizing risk

The microeconomic motivation for this section is the model of Horst and Moreno [12]. We present an overview for completeness. The principal's income, which is exposed to non-hedgeable risk factors, is represented by $W \leq 0$. The latter is a bounded

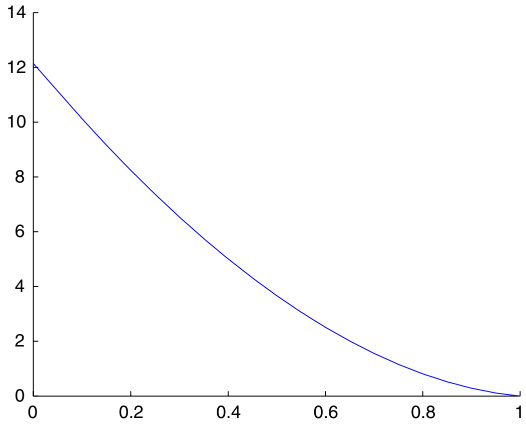


Fig. 8 The approximating function $\overline{v}_{25}$

random variable defined on a standard, non-atomic, probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The principal's goal is to lay off parts of his/her risk with the agents whose preferences are mean-variance. The agent types are indexed by their coefficients of risk aversion, which are assumed to lie $\Theta = [a, 1]$ for some $a > 0$. Notice that the variational problem that arises in this example is structurally different than the previous ones. We have included it to show that our method can be used in more general setting. The principal underwrites call options on her income with type-dependent strikes:

$$X(\theta) = (|W| - K(\theta))^+ \quad \text{with} \quad 0 \leq K(\theta) \leq \|W\|_\infty.$$

If the principal issues the catalogue $\{(X(\theta), \pi(\theta))\}$, he/she receives a cash amount of $\int_\Theta \pi(\theta)d\theta$ and is subject to the additional liability $\int_\Theta X(\theta)d\theta$. He/she evaluates the risk associated with her overall position

$$W + \int_\Theta (\pi(\theta) - X(\theta))d\theta$$

via the "entropic measure" of his/her position, i.e.

$$\rho\left(W + \int_\Theta (X(\theta) - \pi(\theta))d\theta\right),$$

where $\rho(X) = \log(\mathrm{E}[\exp\{-\beta X\}])$ for some $\beta > 0$. The principal's problem is to devise a catalogue as to minimize his/her risk exposure. Namely, he/she chooses a function v and contracts X from the set

$$\{(X, v) \mid v \in \mathcal{C},\ v \leq K_1,\ -\mathrm{Var}[(|W| - K(\theta))^+] = v'(\theta),\ |v'| \leq K_2,\ 0 \leq K(\theta) \leq \|W\|_\infty\},$$

in order to minimize

$$\rho\left(W-\int\limits_{\Theta}\left\{(|W|-F(v'(\theta)))^{+}-\mathbb{E}[(|W|-F(v'(\theta)))^{+}]\right\}d\right)-I(v),$$

where

$$I(v)=\int\limits_{\Theta}\big(\theta v'(\theta)-v(\theta)\big)\,d\theta.$$

We assume the set of states of the World is finite with cardinality m. Each possible state ω_j can occur with probability p_j. The realizations of the principal's wealth are denoted by $W=(W_1,\ldots,W_m)$. Note that p and W are treated as known data. The objective function of our non-linear program is

$$\begin{aligned}F(v,v',K)=\log&\left(\exp\left\{-\sum_{i=1}^{n}W_i p_i+\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}T(K_j-|W_i|)\right)p_i\right.\right.\\&\left.\left.-\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}T(K_j-|W_i|)\right)p_i\right\}\right)+\frac{1}{n}\sum v_i-\theta_i v_i'\end{aligned}$$

where $K=(K_1,\ldots,K_n)$ denotes the vector of type dependent strikes. We denote by ng the total number of constraints. The principal's problem is to find

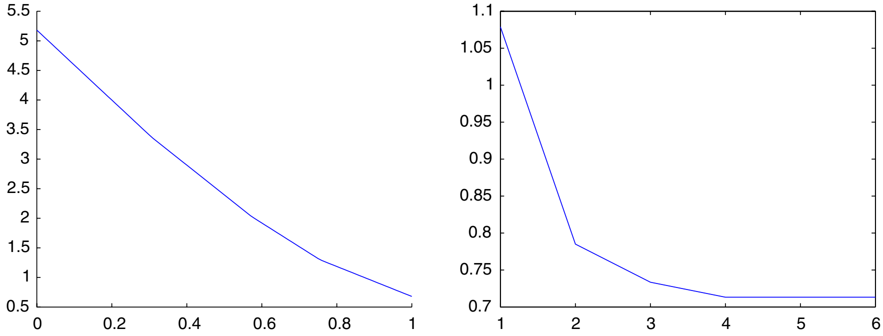
$$\min_{(v,v',K)}F(v,v',K)\quad\text{subject to}\quad G(v,v',K)\le 0,$$

where $G:\mathbb{R}^{3n}\to\mathbb{R}^{ng}$ determines the constraints that keep (v,v',K) within the set of feasible contracts. Let $(1/6,2/6,\ldots,1)$ be the uniformly distributed agent types,

$$\begin{aligned}W&=4*(-2,-1.7,1.4,-.7,-.5,0)\quad\text{and}\\P&=(1/10,1.5/10,2.5/10,2.5/10,1.5/10,1/10).\end{aligned}$$

The principal's initial evaluation of his/her risk is 1.52. Figure 9 shows the plots for the approximating $\overline{v}$ and the strikes. Note that for illustration purposes we have changed the scale for the agent types in the second plot. The interpolates of the approximate to the optimal function $\overline{v}$ and the strikes are given in Table 1. After the exchanges with the agents, the principal's valuation of his/her risk decreases to -3.56.

Remark 4.2 Notice the "bunching" at the bottom.



(a) The optimal function v. **(b)** The type-dependent strikes

Fig. 9 Optimal solution for underwriting call options

Table 1 The optimal function $\overline{v}$ and the strikes

$\overline{v}_1$	4.196344	K_1	1.078869
$\overline{v}_2$	3.234565	K_2	0.785079
$\overline{v}_3$	2.321529	K_3	0.733530
$\overline{v}_4$	1.523532	K_4	0.713309
$\overline{v}_5$	0.745045	K_5	0.713309
$\overline{v}_6$	0.010025	K_6	0.713309

5 Conclusions

In this paper, we have developed a numerical algorithm to estimate the minimizers of variational problems with convexity constraints, with our main motivation stemming from Economics and Finance. Ours is an INTERNAL method, so at each step the approximate minimizers lie within the acceptable set of (convex) functions. This is of particular interest given our microeconomic motivation, where non-convex functions would correspond to non-implementable contracts. Our examples are developed over one or two dimensional sets for illustration reasons, but the algorithm can be implemented in higher dimensions. However, it must be mentioned that, as is the case with the other methods found in the related literature, implementing convexity has a high computational cost which increases geometrically with precision and exponentially with dimension.

A Appendix

In order to prove convergence of our algorithm we make use of the Convex Analysis results contained in this section. We work on Θ, an open and convex subset of $\mathbb{R}^n$ and we use the classical notation $\partial f(\theta)$ to denote the subdifferential of f at θ. Recall that $\partial f(\theta) \neq \emptyset$ for any θ in the interior of the effective domain of f.

Definition A.1 A convex function $f : \Theta \to \mathbb{R}$ is said to be twice A-DIFFERENTIABLE at a point $\theta_0 \in C$ if $\nabla f(\theta_0)$ exists and if there is a symmetric, positive definite matrix $D^2 f(\theta_0)$ (the ALEXANDROV HESSIAN of f at θ_0), such that for any $\epsilon > 0$ there is $\delta > 0$ such that if $\|\theta - \theta_0\| < \delta$, then

$$\sup_{\theta^* \in \partial f(\theta)} \|\theta^* - \nabla f(\theta_0) - D^2 f(t_0)(\theta - \theta_0)\| \leq \epsilon \|\theta - \theta_0\| \tag{20}$$

The following theorem is due to Alexandrov [18]

Theorem A.2 *Let $f : \Theta \to \mathbb{R}$ be convex, then it is almost everywhere twice A-differentiable on Θ. Moreover, if f is A-differentiable at $\theta_0 \in B$, then*

$$\lim_{h \to 0} \frac{f(\theta_0 + h) - f(\theta_0) - \langle \nabla f(\theta_0), h\rangle - \frac{1}{2}\langle D^2 f(\theta_0)h, h\rangle}{\|h\|^2} = 0.$$

Corollary A.3 *Let $f : \Theta \to \mathbb{R}$ be convex. Then the mapping*

$$\theta \to \nabla f(\theta)$$

is well defined and continuous almost everywhere.

Proof Let $\overline{\Theta}$ be the set where f is twice A-differentiable. By definition ∇f is well defined on $\overline{\Theta}$, which is a set of full measure. When restricted to $\overline{\Theta}$, expression (20) can be written as

$$\|\nabla f(\theta) - \nabla f(\theta_0) - D^2 f(t_0)(\theta - \theta_0)\| \leq \epsilon \|\theta - \theta_0\|,$$

which implies continuity of ∇f. □

The following is a well known property of convex functions. We were first made aware of it by Carlier [5], but it probably dates back to Mosco or Joly (see for example [16]).

Proposition A.4 *Let $\Theta \subset \mathbb{R}^n$ be a convex, open set. Assume the sequence of convex functions $\{f_k : \Theta \to \mathbb{R}\}$ converges uniformly to $\bar{f}$, then $\nabla f_k \to \nabla \bar{f}$ almost everywhere on Θ.*

Proof Denote by $D_i f$ the derivative of f in the direction of e_i. The convexity of f_k and $\bar{f}$ implies the existence of a set B, with $\mu(\Theta \backslash B) = 0$ such that the partial derivatives of f_k and $\bar{f}$ exist and are continuous in B. Let $\theta \in B$. To prove that $D_i f_k(\theta) \to D_i \bar{f}(\theta)$, consider η such that $\theta + \eta e_i \in \Theta$. Since f_k is convex

$$\frac{f_k(\theta + he_i) - f_k(\theta)}{h} \geq D_i f_k(\theta) \geq \frac{f_k(\theta - he_i) - f_k(\theta)}{h}$$

for all $0 < h < \eta$. Hence

$$\frac{f_k(\theta+he_i)-f_k(\theta)}{h} - D_i \bar{f}(\theta) \geq D_i f_k(\theta) - D_i \bar{f}(\theta) \geq \frac{f_k(\theta-he_i)-f_k(\theta)}{h} - D_i \bar{f}(\theta).$$

The left-hand side of this inequality is equal to

$$\frac{f_k(\theta + he_i) - \bar{f}(\theta + he_i)}{h} + \frac{\bar{f}(\theta) - f_k(x)}{h} + \frac{\bar{f}(\theta + he_i) - \bar{f}(\theta)}{h} - D_i \bar{f}(\theta).$$

For $\epsilon > 0$ let $0 < \delta < \eta$ be such that

$$\left| \frac{\bar{f}(\theta + he_i) - \bar{f}(\theta)}{h} - D_i \bar{f}(\theta) \right| < \epsilon \tag{21}$$

for $|h| \leq \delta$. Let $N \in \mathbb{N}$ be such that

$$-\epsilon\delta \leq f_n(\theta) - \overline{f}(\theta) \leq \epsilon\delta,$$

$n \geq N$. Hence, taking $h = \delta$ (notice that Eq. 21 still holds in this case), we have that for all $n \geq N$,

$$\frac{f_n(\theta + he_i) - \overline{f}(\theta + he_i)}{h} \leq \epsilon \quad \text{and} \quad \frac{\overline{f}(\theta) - f_n(\theta)}{h} \leq \epsilon.$$

Hence,

$$3\epsilon \geq D_i f_n(\theta) - D_i \overline{f}(\theta)$$

for all $x \in B$. The same argument shows that

$$-3\epsilon \leq D_i f_n(\theta) - D_i \overline{f}(\theta)$$

for all $n \geq N$ and all $x \in B$, which concludes the proof. □

Proposition A.5 *Let $\Theta \subset \mathbb{R}^n$ be a convex, compact set and let $g : U \to \mathbb{R}$ be a convex function such that for all $\theta \in \Theta$, the subdifferentials $\partial g(\theta)$ are contained in Q for some compact set Q. Then there exists $\{g_j : \Theta \to \mathbb{R}\}$ such that $g_j \in \mathcal{C}^1(\Theta)$ and $g_j \to g$ uniformly on Θ.*

Proof Fix $\delta > 0$ and define

$$\Theta_\delta := \{(1+\delta)x \mid x \in \Theta\}.$$

Extend g to be defined on Θ_δ. Define

$$\Psi(\theta) := \begin{cases} Ke^{-\frac{1}{1-\|\theta\|^2}}, & \|\theta\| < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Here K is chosen so that $\int_{\mathbb{R}^n} \Psi(\theta)d\theta = 1$. For each $\epsilon > 0$ define the *mollifier* $\Psi_\epsilon(\theta) := \epsilon^{-n}\Psi(\theta/\epsilon)$ (for a discussion on the properties of mollifiers see, for example [11]). The functions

$$h_\epsilon := g * \Psi_\epsilon$$

are convex, smooth and they converge uniformly to g on U as long as ϵ is small enough so that

$$U_\epsilon := \{\theta \in \Theta_\delta \mid d(x, \partial\Theta_\delta) > \epsilon\}$$

is contained in Θ. Let $n \in \mathbb{N}$ be such that $T_{1/n} \subset \Theta$, then the sequence $\{g_j := h_{1/j}\}$ has the required properties. □

Lemma A.6 *Consider $\phi(\theta, z, p) \in \mathcal{C}^1(\Theta \times \mathbb{R} \times Q \to \mathbb{R})$, where $\Theta = [a, b]^n$ and Q is a compact convex subset of $\mathbb{R}^n$. Let $\{f_k : \Theta \to \mathbb{R}\}$ be a family of convex functions such that $\partial f_K(\theta) \subset Q$ for all $\theta \in \Theta$, and whose uniform limit is $\overline{f}$. Let Σ_k be the uniform partition of Θ consisting of k^n cubes of volume $\mu(\Sigma_k) := (\frac{b-a}{k})^n$. Denote by σ_j^k, $1 \le j \le k^n$, be the elements of Σ_k and let*

$$\mu(\Sigma_k) \sum_{i=1}^{k^n} \phi\left(\theta_j^k, f_k(\theta_j^k), \nabla f_k(\theta_j^k)\right)$$

be the corresponding Riemann sum approximating $\int_\Theta \phi(x, f_k(\theta), \nabla f_k(\theta))d\theta$, where $\theta_j^k \in \sigma_j^k$ and $\sigma_j^k \in \Sigma_k$. Then for any $\epsilon > 0$ there is $K \in \mathbb{N}$ such that

$$\left| \int_\Theta \phi(\theta, f_k(\theta), \nabla f_k(\theta))d\theta - \mu(\Sigma_k) \sum_{i=1}^{k^n} \phi(\theta_j^k, f_k(\theta_j^k), \nabla f_k(\theta_j^k)) \right| \le \epsilon \tag{22}$$

for any $k \ge K$.

Proof By Lemma A.5, for each f_k there exists a sequence of continuously differentiable convex functions $\{g_j^k\}$ such that

$$g_j^k \to f_k \quad \text{uniformly.}$$

Let h_k be the first element in $\{g_j^k\}$ such that $\|h_k - f_k\| \le \frac{1}{k}$ and $\|\nabla h_k(\theta) - \nabla f_k(\theta)\| \le \frac{1}{k}$ for all $\theta \in \Theta$ where ∇f_k is continuous. Then $h_k \to \overline{f}$ uniformly, and by Lemma A.4 we have that $\nabla h_k(\theta) \to \nabla \overline{f}(\theta)$ a.e. It follows from Egoroff's theorem that for every $n \in \mathbb{N}$ there exists a set $\Lambda_n \subset \Theta$ such that:

$$\mu(\Theta \backslash \Lambda_n) < 1/n \quad \text{and} \quad \nabla h_k \to \nabla \overline{f} \quad \text{uniformly on} \quad \Lambda_n.$$

Let $\mathcal{X}_{\sigma_j^k}(\cdot)$ be the indicator function of σ_j^k and define

$$g_k(\theta) := \phi(\theta, h_k(\theta), \nabla k_k(\theta)) - \sum_{j=1}^{k^n} \mathcal{X}_{\sigma_j^k}(\theta)\phi(\theta_j^k, h_k(\theta_j^k), \nabla h_k(\theta_j^k)).$$

Fix n, consider $\theta_0 \in \Lambda_n$ and let $\{\theta_0^k\}$ be the sequence of θ_j^k's converging to θ_0 as the partition is refined. By uniform convergence, $\nabla \overline{f}$ is continuous on Λ_n, hence

$$\lim_{k\to\infty} h_k(\theta_0^k) = \overline{f}(\theta_0) \quad \text{and} \quad \lim_{k\to\infty} \nabla h_k(\theta_0^k) = \nabla \overline{f}(\theta_0). \tag{23}$$

It follows from (23) and the continuity of ϕ that $g_k \to 0$ almost everywhere on Λ_n. Notice that as a consequence of the compactness of Θ and Q and the definition of h_k we have

$$\|\phi(\theta, f_k(\theta), \nabla f_k(\theta))\| \leq K_1, \quad \text{for all} \quad \theta \in \Theta \quad \text{and some} \quad K_1 > 0,$$

and

$$\left| g_k(\theta) - \left(\phi(\theta, f_k(\theta), \nabla f_k(\theta)) - \sum_{j=1}^{k^n} \mathcal{X}_{\sigma_j^k}(\theta)\phi(\theta_j^k, f_k(\theta_j^k), \nabla f_k(\theta_j^k)) \right) \right| \leq \frac{K_2}{k}$$

for some $K_2 > 0$ and all $\theta \in \Theta$ where ∇f_k is continuous. Therefore

$$\begin{aligned} &\left| \int_\Theta \phi(\theta, f_k(\theta), \nabla f_k(\theta)) d\theta - \|\sigma_j^k\| \sum_{i=1}^{k^n} \phi(\theta_j^k, f_k(\theta_j^k), \nabla f_k(\theta_j^k)) \right| \\ &\quad \leq \frac{K_2 \|\Theta\|}{k} + \left| \int_\Theta g_k(\theta) d\theta \right|. \end{aligned} \tag{24}$$

By Lebesgue Dominated Convergence

$$\lim_{k\to\infty} \int_{\Lambda_n} g_k(\theta) d\theta = 0,$$

moreover, the definition of Λ_n implies

$$\int_{\Theta \setminus \Lambda_n} g_k(\theta) d\theta \leq \frac{2K_1}{n}.$$

Given $\epsilon > 0$ take $n \in \mathbb{N}$ such that $\frac{2K_1}{n} \leq \frac{\epsilon}{2}$ and K such that

$$\frac{K_2\|\Theta\|}{K} + \left| \int\limits_{\Lambda_n} g_K(\theta)d\theta \right| \leq \frac{\epsilon}{2}.$$

Then Eq. (22) holds for all $k \geq K$. □

References

1. Anguilera, N.E., Morin, P.: Approximating optimization problems over convex functions. Numer. Math. **111**, 1–34 (2008)
2. Armstrong, M.: Multiproduct nonlinear pricing. Econometrica **64**, 51–75 (1996)
3. Bolton, P., Dewatripoint, M.: Contract Theory. MIT Press, Cambridge (2005)
4. Brock, F., Ferone, V., Kawohl, B.: A symmetry problem in the calculus of variations. Calc. Var. Partial Differ. Equ. **4**, 593–599 (1996)
5. Carlier, G.: Calculus of variations with convexity constraint. J. Nonlinear Convex Anal. **3**(2), 125–143 (2002)
6. Carlier, G.: Duality and existence for a class of mass transportation problems and economic applications. Adv. Mathe. Econ. **5**, 1–21 (2003)
7. Carlier, G., Ekeland, I., Touzi, N.: Optimal derivatives design for mean-variance agents under adverse selection. Math. Finan. Econ. **1**, 57–80 (2007)
8. Carlier, G., Lachand-Robert, T., Maury, B.: A numerical approach to variational problems subject to convexity constraints. Numer. Math. **88**, 299–318 (2001)
9. Choné, P., Le Meur, H.V.J.: Non-convergence result for conformal approximation of variational problems subject to a convexity constraint. Numer. Funct. Anal. Optim. **22**(5), 529–547 (2001)
10. Ekeland, I., Témam, R.: Convex analysis and variational problems. Classics in Applied Mathematics, vol. 28. SIAM (1976)
11. Evans, L.: Partial differential equations. Graduate Studies in Mathematics, vol. 19. American Mathematical Society, Providence (2002)
12. Horst, U., Moreno-Bromberg, S.: Risk minimization and optimal derivative design in a principal agent game. Math. Finan. Econ. **2**(1), 1–27 (2008)
13. Kawohl, B., Schwab, C.: Convergent finite elements for a class of nonconvex variational problems. IMA J. Numer. Anal. **18**, 133–149 (1998)
14. Lachand-Robert, T., Oudet, É.: Minimizing within convex bodies using a convex hull method. Siam J. Optim. **16**(2), 368–379 (2005)
15. Lachand-Robert, T., Peletier, A.: Extremal points of a functional on the set of convex functions. In: Proceedings of the American Mathematical Society, vol. 127–136, pp. 1723–1727 (1999)
16. Mosco, U.: Convergence of convex sets and of solutions of variational inequalities. Adv. Math. **3**, 510–585 (1969)
17. Mussa, M., Rosen, S.: Monopoly and product quality. J. Econ. Theory **18**, 301–317 (1978)
18. Nicolescu, C., Persson, L.-E.: Convex Functions and their Applications. CMS Books in Mathematics. Springer Science+Business Media, Heidelberg (2006)
19. Rochet, J.-C., Choné, P.: Ironing, sweeping and multidimensional screening. Econometrica **66**, 783–826 (1988)