Approximating optimization problems over convex functions

Néstor E. Aguilera · Pedro Morin

Abstract   Many problems of theoretical and practical interest involve finding an optimum over a family of convex functions. For instance, finding the projection on the convex functions in $H^k(\Omega)$, and optimizing functionals arising from some problems in economics. In the continuous setting and assuming smoothness, the convexity constraints may be given locally by asking the Hessian matrix to be positive semidefinite, but in making discrete approximations two difficulties arise: the continuous solutions may be not smooth, and functions with positive semidefinite discrete Hessian need not be convex in a discrete sense. Previous work has concentrated on non-local descriptions of convexity, making the number of constraints to grow super-linearly with the number of nodes even in dimension 2, and these descriptions are very difficult to extend to higher dimensions. In this paper we propose a finite difference approximation using positive semidefinite programs and discrete Hessians, and prove convergence under very general conditions, even when the continuous solution is not smooth, working on any dimension, and requiring a linear number of constraints in the number of nodes. Using semidefinite programming codes, we show concrete examples of approximations to problems in two and three dimensions.

Mathematics Subject Classification (2000)   52B55 · 65D15 · 90C90

N. E. Aguilera · P. Morin (✉)
Consejo Nacional de Investigaciones Científicas y Técnicas
and Universidad Nacional del Litoral, Güemes 3450, S3000GLN Santa Fe, Argentina
e-mail: pmorin@santafe-conicet.gov.ar

N. E. Aguilera
e-mail: aguilera@santafe-conicet.gov.ar

1 Introduction

Convex and concave functions appear naturally in many branches of science such as biology (growth), medicine (dose-response), or economics (utility, production or costs), spurring in turn the interest of other areas, for example, statistics. It is no surprise, then, that many problems of theoretical and practical interest involve the optimization of a functional over a family of convex functions.

A particularly important case is when the functional to be minimized is the norm of some normed function space V, i.e., given $f \in V$ find

$$\min_{u \in \mathcal{C}} \|u - f\|_V,$$

where $\mathcal{C}$ is a family of convex functions in V. Typical choices are the spaces $L^2(\Omega)$ or $H^1(\Omega)$, and, in general, the $W^{k,p}(\Omega)$ spaces, where Ω is a convex domain in $\mathbb{R}^d$.

Sometimes the convexity is a reasonable shape assumption on the model, which could be replaced by or added to other shape constraints such as radial symmetry, harmonicity or upper and lower bounds. This is the case, for example, of Newton's problem of minimal resistance (see, e.g. [3,4,12,13]).

More surprisingly perhaps, the convexity constraint may be a consequence of the model, as in the design of some mechanisms in economics [14,17]. Actually, our interest in the subject arose from one of these problems, which we will call the *monopolist problem*, and is described in some detail in Sect. 6.1. In this problem we wish to find

$$\max_{u \in \mathcal{C}} \int\limits_Q (\nabla u(x) \cdot x - u(x)) \ f(x) \, \mathrm{d}x, \tag{1.1}$$

where $Q = [0, 1]^d$, f is a non-negative probability density function over Q, and $\mathcal{C}$ is a family of convex functions on Q with some further properties.

The monopolist problem is numerically very challenging since, unlike the problems coming from physics, the space dimension d may be much higher than 2 or 3.

There is a big qualitative jump going from one to more dimensions when dealing with convexity constraints. This can be appreciated readily by looking at the statistics literature, where the one dimensional convex regression or density problem has been considered both theoretically and numerically, and using different measuring functionals (see, e.g. [1,8,9,15]), but very little has been done numerically in two or more dimensions. In this regard, it is worth mentioning the work by Shih, Chen and Kim [18], who use appropriate splines in the MARS multivariate adaptive regression splines (MARS) algorithm.

One of the main difficulties in obtaining discrete approximations to optimization problems over convex functions lies in giving a local and finite description of convex functions if $d > 1$. Though this could be done for smooth functions of continuous variables by asking the Hessian matrix to be positive semidefinite at all points, there is no similar characterization for discrete functions on meshes. As a matter of fact, we show in Examples 2.3 and 2.4 that this cannot be done easily.

It goes without saying that there is a lot of work done on convex functions of continuous variables, as exemplified by the classical book by Rockafellar [19]. Traditionally, this work leans more on properties satisfied by convex functions rather than on properties that imply convexity. In particular, very little has been done on local properties guaranteeing convexity.

For discrete variables, there is a rather large work done by the discrete mathematics community for fixed lattices, and a number of definitions for discrete convex functions have been proposed (see, e.g. [16]). Again, usually these definitions are non-local.

As far as we know, there exists very few literature on optimization problems on convex functions in dimensions two or more, either theoretically or numerically. Besides those already mentioned (and the references therein), let us cite three more which are prominent in the context of our work.

- Carlier and Lachand-Robert [5] obtained the C^1 regularity of a variant of the monopolist problem, substituting the functional in (1.1) for

$$\int_\Omega \left(-\frac{1}{2} |\nabla u|^2 + x \cdot \nabla u - u(x) \right) f(x) \, dx,$$

  under some restrictions on the domain Ω and the density f. They obtained also C^1 regularity for convex minimizers of functionals of the form

$$\int_\Omega (A(x, u)\nabla u(x) \cdot \nabla u(x) + f(x)u(x)) \, dx.$$

- Carlier, Lachand-Robert and Maury [6] proposed a numerical scheme for minimizers of functionals of the type

$$\int_\Omega j(x, u(x), \nabla u(x)) \, dx,$$

  on closed convex subsets of convex functions of $H^1(\Omega)$ or $L^2(\Omega)$, where j is a quadratic function of u and ∇u, and $\Omega \subset \mathbb{R}^2$. We will discuss their approach in the next section [after the inequalities (2.5)].
  As Carlier et al. [6] point out, their work encompasses the problem of finding

$$\min \int_\Omega |u - f|^2 \, dx \quad \text{subject to} \quad u \in L^2(\Omega), \ u \text{ convex}, \ u \leq f,$$

  for given $f \in L^2(\Omega)$, i.e., a L^2-norm projection, and this problem is equivalent to that of finding the convex envelope f^{**} of f. Thus, minimizing over convex functions and finding the convex envelope of a function are two quite related tasks.
- Being one of the main problems in computational geometry, there are a number of well established codes for finding the convex hull of a set of points in $\mathbb{R}^d$, which

are very efficient in low dimensions. Hence, it is natural to try to use these codes to approximate optimization problems on convex functions, an approach which Lachand-Robert and Oudet [11] applied to several problems.

It would be very interesting to see whether these ideas could be carried over, since the convex hull codes are quite fast in low dimensions.

Our approach, based on semidefinite programming, takes a different direction from those mentioned.

Let us recall that a semidefinite program is an optimization problem of the form

$$\min c \cdot x$$

$$\text{subject to}$$
$$x_1 A_1 + \cdots + x_n A_n - A_0 \succeq 0, \qquad (1.2)$$
$$x \in \mathbb{R}^n,$$

where $c \in \mathbb{R}^n$, $A_0, A_1, \ldots, A_n$ are symmetric $m \times m$ matrices, and $A \succeq 0$ indicates that the symmetric matrix A is positive semidefinite. By letting the matrices A_i be diagonal, we see that the program (1.2) is a generalization of linear programming (and includes it strictly). Thus, in a semidefinite program the constraints can be a mixture of linear inequalities and positive semidefinite requirements.

In this paper we give a theoretical framework for approximating many optimization problems on convex functions using a finite differences scheme which imposes a positive semidefinite constraint on a discretization of the Hessian matrix.

Although not linear, our approach seems very natural and has many advantages. Being of a local nature, the number of constraints grows only linearly with the number of nodes, and it works for any dimension of the underlying space. Notwithstanding the already mentioned counterexamples (2.3 and 2.4) of the relation between convexity and positive semidefinite Hessian, we will show that for many problems we obtain convergence to a continuous optimum. This convergence holds even for non-smooth optima, as those arising when projecting in the L^∞ norm or in some problems of the type given by Eq. (1.1).

In practice, our definition can be used to advantage by using the existing efficient semidefinite codes and, furthermore, it is very simple to program in higher dimensions.

As a final remark, we think that our approach might be a first step to deal with other problems where convexity is a consequence of the model, such as transportation problems where the cost is quadratic, i.e., for solving the Monge–Ampère equations, a prime example of fully non-linear differential equations.

This article is organized as follows.

In Sect. 2 we summarize some techniques that could be used to deal with the numerical approximation of optimal convex functions, showing their strengths and weaknesses, and introduce the discrete Hessian.

In Sect. 3 we give different characterizations of smooth convex functions of continuous variables, which allow us to extend the definition of the Hessian to C^1 functions in terms of averages.

The heart of the paper is Sect. 4: in order to have a good definition of discrete convexity, we need to show that any convex function of continuous variables may be

approximated by its discrete counterparts, and conversely, that a converging sequence (in a suitable norm) of discrete convex functions will do so to a convex function of continuous variables.

We show in this section that the approximation of the discrete functions to the limit u is uniform in u and its derivatives if u is smooth, and uniform in u over compact subsets of Ω, if u is merely convex and bounded (and hence continuous). Although our definition is stated using the Hessian, if a sequence of discrete functions converges to a function which is not C^2 or even C^1, still convexity of the limit function is guaranteed.

In Sect. 5 we pose a general structure for optimization problems on convex functions which fits the results of the previous sections, and may be applied to several important problems. In addition, error bounds may be obtained if the continuous optimal solution is smooth.

The results of Sects. 3–5 are somewhat independent of semidefinite programming. However, as explained in Sect. 2, for many problems of interest the functional and constraints involved may be expressed as a semidefinite program, and we present in Sect. 6 numerical examples showing that the current codes make it feasible to use our scheme on them.

We begin Sect. 6 by considering the monopolist problem in two and three dimensions, comparing our results to the analytical solution when $f = 1$ in (1.1). We continue by showing some norm projections, exploring the behavior under different functionals of an example given by Carlier et al. [6]. Moreover, we exhibit an explicit example in which the L^∞ projection is not unique. We finish the section on numerical examples by showing how our scheme could be used to fit a (discrete) convex function to noisy data given on the nodes of a regular mesh, using the L^2 norm (i.e., a least squares approach), and also the L^1 and L^∞ norms, which are not often seen.

2 Dealing with convex functions

In this section we present some ideas and techniques that could be used to approximate numerically an optimal convex function. For the sake of simplicity, we will work on the unit d-dimensional cube in $\mathbb{R}^d$, $Q = [0, 1]^d$.

If $d = 1$ and

$$x_0 = 0 < x_1 < \cdots < x_n = 1, \tag{2.1}$$

is a discretization of $[0, 1]$, then a convex function u defined on $[0, 1]$ satisfies the inequalities

$$\frac{u(x_i) - u(x_{i-1})}{x_i - x_{i-1}} \leq \frac{u(x_{i+1}) - u(x_i)}{x_{i+1} - x_i} \quad \text{for } i = 1, \ldots, n-1. \tag{2.2}$$

Conversely, if a function u defined only in the mesh points $\{x_0, \ldots, x_n\}$ satisfies the inequalities (2.2), we can always extend it linearly in the intervals $[x_{i-1}, x_i]$, $i = 1, \ldots, n$, so that the resulting piecewise linear function will be convex on $[0, 1]$.

Thus, the inequalities (2.2) may be used to define discrete convexity in one dimension. It is clear that—via the piecewise linear extension—we can approximate any

convex function in [0, 1]. On the other hand, as we show in Lemma 2.1, under some adequate assumptions if a sequence of discrete convex functions [satisfying (2.2) for each corresponding subdivision] converges pointwise, then the limit defines a unique convex function in [0, 1].

Let h be given, $0 < h < 1/2$, and suppose $\mathcal{M} = \{x_0, \ldots, x_n\}$ is a subdivision of [0, 1] satisfying (2.1) and such that

$$\max_{1 \leq i \leq n} |x_i - x_{i-1}| \leq h. \tag{2.3}$$

Let u be defined on $\mathcal{M}$ and satisfy (2.2). For a given ε, $h < \varepsilon < 1/2$, let us consider

$$\underline{x} = \max \{x \in \mathcal{M} : x \leq \varepsilon\} \quad \text{and} \quad \overline{x} = \min \{x \in \mathcal{M} : x \geq 1 - \varepsilon\}.$$

From (2.2) we see that

$$-\frac{|u(\underline{x}) - u(0)|}{\varepsilon} \leq \frac{u(\underline{x}) - u(0)}{\underline{x}} \leq \frac{u(x_{i+1}) - u(x_i)}{x_{i+1} - x_i} \leq \frac{u(1) - u(\overline{x})}{1 - \overline{x}} \leq \frac{|u(1) - u(\overline{x})|}{\varepsilon}$$

for all $x_i, x_{i+1} \in \mathcal{M}$ such that $\underline{x} \leq x_i, x_{i+1} \leq \overline{x}$. Therefore, if

$$|u(x_i)| \leq M \quad \text{for all } x_i \in \mathcal{M},$$

we have

$$\underline{C} \leq \frac{u(x_{i+1}) - u(x_i)}{x_{i+1} - x_i} \leq \overline{C} \quad \text{for all } \varepsilon \leq x_i, x_{i+1} \leq 1 - \varepsilon, \tag{2.4}$$

where $\underline{C}$ and $\overline{C}$ are constants depending only on ε and M (but not on u or $\mathcal{M}$). In other words, if u is extended as a piecewise linear function to all of [0, 1], the resulting function is uniformly Lipschitz on compact subsets of (0, 1).

Hence,

Lemma 2.1 *Let $(\mathcal{M}_h)_h$ be a sequence of meshes in* [0, 1], *$h \downarrow 0$, satisfying* (2.3) *and satisfies* (2.2) *for each h, and such that $\mathcal{M}_h \subset \mathcal{M}_{h'}$ for $h > h'$. Suppose u_h is defined in $\mathcal{M}_h$ for each h, and the sequence $(u_h)_h$ is such that for all $x \in \cup_h \mathcal{M}_h$, $\lim_h u_h(x)$ exists and is finite, and let*

$$u(x) = \lim_h u_h(x) \quad \text{for } x \in \bigcup_h \mathcal{M}_h.$$

Then u may be extended to all of [0, 1] *as a continuous convex function in a unique way. Moreover, the convergence of u_h (extended piecewise linearly) to u is uniform in compact subsets of* (0, 1).

Also, by the Arzela–Ascoli theorem,

Lemma 2.2 *Let $(\mathcal{M}_h)_h$ be a sequence of meshes as in the previous lemma. Suppose $M > 0$ is given, and that for each h a function u_h is defined in $\mathcal{M}_h$ and satisfies (2.2) so that*

$$|u_h(x)| \leq M \quad \text{for all } x \in \mathcal{M}_h \text{ and all } h,$$

and assume u_h is extended to all of $[0, 1]$ piecewise linearly.

Then, there exists a subsequence of $(u_{h'})$ of (u_h) converging to a continuous convex function u defined on $[0, 1]$. Moreover, the convergence of the subsequence $(u_{h'})$ to u is uniform in compact subsets of $(0, 1)$.

Thus, from a theoretical point of view, the discrete functions satisfying (2.2) are well understood.

From the numerical point of view, when used in a discretization of an optimization problem, the constraints coming from the inequalities (2.2) are linear, and solving the resulting discrete optimization problem with them is usually not much harder than solving it without them.

Hence, the case $d = 1$ poses no major trouble.

For $d > 1$ it will be more convenient to work only with regular meshes (or grids) on $Q = [0, 1]^d$. Thus, for fixed $h > 0$ ($h = 1/n$ for some $n \in \mathbb{N}$), the mesh $\mathcal{M}_h$ will consist of all points $x \in \mathbb{R}^d \cap Q$ of the form $x = hz$ with $z \in \mathbb{Z}^d$. Denoting by $\mathring{Q} = Q \setminus \partial Q$ the interior of Q, we set

$$\mathring{\mathcal{M}}_h = \mathcal{M}_h \cap \mathring{Q}, \qquad \partial \mathcal{M}_h = \mathcal{M}_h \cap \partial Q,$$

and denote by $\mathcal{U}_h$ the set of real valued functions defined on $\mathcal{M}_h$.

A first simple idea to extend the inequalities (2.2) to more dimensions, is to consider the set of functions $u_h \in \mathcal{U}_h$ satisfying the convexity constraints

$$u_h(x + h\mathbf{e}_i) + u_h(x - h\mathbf{e}_i) \geq 2\,u_h(x) \quad \text{for all } x \in \mathring{\mathcal{M}}_h \quad \text{and} \quad i = 1, \ldots, d,$$

where $\mathbf{e}_i$ denotes the ith vector of the canonical basis of $\mathbb{R}^d$.

This set of discrete functions is very appealing because the convexity is modelled by using only $O(\text{number of mesh points})$ linear constraints, as in the case $d = 1$. As we have seen, this approach is exact in one dimension, and gives satisfactory results in many cases in more dimensions (in particular for some of the specific examples treated later in Sect. 6.1), but there is no guarantee of convexity in the limit function if $d > 1$. In fact, by taking the interpolant, with this set we can certainly approximate any function of continuous variables which is convex, but we can also approximate other functions. For example, we may approximate $u(x_1, x_2) = x_1 x_2$ which is linear—and thus convex—in each coordinate direction, but not convex in $Q \subset \mathbb{R}^2$.

This shows that the definition of discrete convexity should be done with more care: when $d > 1$, we have to take into account every possible direction.

It is reasonable, then, to say that a discrete function is convex if

$$u_h(x + y) + u_h(x - y) \geq 2u_h(x) \quad \text{for all } x, y \text{ such that } x \pm y \in \mathcal{M}_h. \quad (2.5)$$

Since as h goes to 0 the possible directions become dense, it is possible (under some conditions) to regain convexity in the limit as we had for one dimension.

In a way, this is the approach followed by Carlier et al. [6]. In a two dimensional setting, they consider discrete convex functions which are restrictions to a mesh of convex functions of continuous variables, and show that this definition is equivalent to an intrinsic one, stated only in terms of the value of the function at the grid points, similar to (2.5). The problem with this description is that it is non-local, and the number of constraints needed in two dimensions (after pruning) reportedly grows approximately as $N^{1.8}$, where N is the number of nodes in the grid. Moreover, this approach is very difficult to extend to higher dimensions.

In order to keep the definition of discrete convexity local, another possibility is to consider discretizations of the Hessian matrix, as we do in this work.

For $u \in \mathcal{U}_h$ and $x \in \mathcal{M}_h$, we define the (forward) first order finite differences by

$$\Delta_{h,i}\, u(x) = \frac{u(x + h\, \mathbf{e}_i) - u(x)}{h}, \qquad i = 1, \ldots, d,$$

and the second order finite differences by

$$\Delta^2_{h,ii}\, u(x) = \frac{u(x + h\, \mathbf{e}_i) - 2\, u(x) + u(x - h\, \mathbf{e}_i)}{h^2}, \qquad i = 1, \ldots, d,$$

or, if $i \neq j$, by

$$\begin{aligned}
\Delta^2_{h,ij}\, u(x) = \frac{1}{4\, h^2}\, \big(& u(x + h\, \mathbf{e}_i + h\, \mathbf{e}_j) - u(x - h\, \mathbf{e}_i + h\, \mathbf{e}_j) \\
& - u(x + h\, \mathbf{e}_i - h\, \mathbf{e}_j) + u(x - h\, \mathbf{e}_i - h\, \mathbf{e}_j) \big).
\end{aligned}$$

Clearly, not all of these finite differences are defined for points in $\partial \mathcal{M}_h$. Therefore, when mentioning $\Delta_{h,i}\, u(x)$ or $\Delta^2_{h,ij}\, u(x)$ for $x \in \mathcal{M}_h$, we will implicitly assume that x and i (and eventually j) are such that the corresponding finite difference is well defined at x.

Finally, we define the discrete Hessian of $u \in \mathcal{U}_h$ at $x \in \mathring{\mathcal{M}}_h$, as the symmetric matrix $H_h u(x) \in \mathbb{R}^{d \times d}$ whose i, j entry is $\Delta^2_{h,ij}\, u(x)$.

We are faced now with two issues:

1. How can a discrete optimization problem with positive semidefinite discrete Hessian constraints be solved?
2. Once we found a discrete solution, how does it approximate the continuous solution?

As mentioned in the introduction, for the first issue we will use the semidefinite programming model (1.2), where the objective is linear and the constraints are positive semidefinite. Thus, if the objective of the continuous problem is not linear, we must rewrite it as a constraint. However, it is worth mentioning that not every functional may be readily modelled as a positive semidefinite constraint, an example of such functionals being

$$\int_{\Omega} \frac{1}{1 + |\nabla u|^2} \, dx,$$

arising in Newton's problem of minimal resistance.

Although this method may not be used for *all* functionals, it can be used in *many* cases as we now illustrate. The interested reader is referred to the article by Vandenberghe and Boyd [20] for other possibilities.

In what follows we associate with each mesh node $P_k, k = 1, \ldots, N$, the unknown value u_k, the discrete Hessian $H_h(P_k)$, and a d-dimensional cube Q_k centered at P_k of measure $|Q_k|$, so that $\sum_k |Q_k| = |Q|$.

L^1 **projection.** The program

$$\min \int_Q |u - f| \, dx = \min \sum_k \int_{Q_k} |u - f| \, dx,$$

subject to u convex, is modeled by adding N more unknowns $t_1, \ldots, t_N$ (for a total of $2N$) as

$$\min \sum_k t_k$$

subject to

$$\int_{Q_k} |u - f| \, dx \leq t_k, \quad k = 1, \ldots, N,$$

$$H_h(P_k) \succeq 0, \quad k = 1, \ldots, N.$$

In turn, the constraints of the form

$$\int_{Q_k} |u - f| \, dx \leq t_k,$$

are lumped as $|u_k - f(P_k)| \, |Q_k| \leq t_k$, and finally modeled as the linear inequalities

$$-u_k |Q_k| + t_k \geq -f(P_k) |Q_k|,$$

$$u_k |Q_k| + t_k \geq f(P_k) |Q_k|.$$

L^2 **projection.** This is analogous to the L^1 projection:

$$\min \sum_k t_k$$

subject to

$$(u_k - f_k)^2 |Q_k| \leq t_k, \quad k = 1, \ldots, N$$

$$H_h(P_k) \succeq 0, \quad k = 1, \ldots, N,$$

and the constraints of the form $(u_k - f_k)^2 |Q_k| \leq t_k$ are written in the form

$$\begin{bmatrix} t_k & (x_k - f_k)\sqrt{|Q_k|} \\ (x_k - f_k)\sqrt{|Q_k|} & 1 \end{bmatrix} \succeq 0.$$

The H^1 projection is handled in a similar way.

L^∞ **projection.** This is analogous to the other projections, except that we only need to add one more variable t (instead of N), obtaining the discrete program

$$\min t$$

subject to
$$-u_k + t \geq -f(P_k), \quad k = 1, \ldots, N,$$
$$u_k + t \geq f(P_k), \quad k = 1, \ldots, N,$$
$$H_h(P_k) \succeq 0, \quad k = 1, \ldots, N.$$

The next issue is to see how well the discrete solutions of the semidefinite program approximate the continuous convex solution.

Our first hope is that $H_h u(x) \succeq 0$ will be equivalent to some version of convexity of discrete functions, for example to the one given by the inequalities (2.5). The next examples show that this is not true. The first one shows that a non-convex function may have a positive semidefinite discrete Hessian, and the second one shows that the discrete Hessian may not be positive semidefinite for some convex functions.

Example 2.3 Let us consider $d = 2$, and u defined on the 2-dimensional 2×2 grid $\mathcal{M}_{1/2}$ by

$$\begin{array}{lll} u(0, 1) = 1, & u(1/2, 1) = 1, & u(1, 1) = 1, \\ u(0, 1/2) = 1/2, & u(1/2, 1/2) = 5/8, & u(1, 1/2) = 1, \\ u(0, 0) = 0, & u(1/2, 0) = 1/2, & u(1, 0) = 1. \end{array}$$

We may check that (with $h = 1/2$),

$$H_h u(1/2, 1/2) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

whose eigenvalues are 2 and 0 [with eigenvectors $(1, -1)$ and $(1, 1)$], and is thus positive semidefinite.

However, the restriction to the diagonal $x_1 = x_2$ is not convex:

$$\frac{u(0, 0) + u(1, 1)}{2} = \frac{1}{2} < \frac{5}{8} = u(1/2, 1/2).$$

Example 2.4 Consider the same grid of the previous example and u defined by

$$\begin{array}{lll} u(0, 1) = 8/15, & u(1/2, 1) = 8/15, & u(1, 1) = 1, \\ u(0, 1/2) = 1/30, & u(1/2, 1/2) = 1/2, & u(1, 1/2) = 1, \\ u(0, 0) = 0, & u(1/2, 0) = 1/2, & u(1, 0) = 1. \end{array}$$

We see that u is the restriction to $\mathcal{M}_h$ of the convex function

$$\max \left\{ x_1, \frac{14x_1 + x_2}{15}, x_2 - \frac{7}{15} \right\},$$

but

$$H_h u(1/2, 1/2) = \frac{1}{15} \begin{bmatrix} 2 & -23 \\ -23 & 2 \end{bmatrix},$$

which has eigenvalues $5/3$ and $-7/5$, and hence is not positive semidefinite.

Notwithstanding these examples, in Sect. 4 we will show that, under suitable conditions, by asking for positive semidefinite discrete Hessians we may conveniently approximate continuous convex functions, and obtain variants of Lemmas 2.1 and 2.2.

However, we will need some previous results which are tackled in the following section.

3 Convex functions of continuous variables

In this section we deal with variants of the Hessian matrix, initially defined for C^2 functions, which allow us to characterize convexity for C^1 functions locally.

Let us start by introducing some more precise notation:

- The distance from a point x to a set A will be denoted by $\mathrm{dist}(x, A)$.
- If $x \in \mathbb{R}^d$ and $h > 0$, $Q_h(x) = \{y \in \mathbb{R}^d : |y_i - x_i| \le h \text{ for all } i = 1, \dots, d\}$.
- The gradient of u is denoted by $\nabla u = (\partial_1 u, \dots, \partial_d u)$, and we write ∂_{ij}^2 for $\partial_i \partial_j$.
- If $\Omega \subset \mathbb{R}^d$ is an open set, we denote by $C^k(\Omega)$ the set of functions having all derivatives up to order k continuous on Ω, and by $C^k(\overline{\Omega})$ the set of functions having all derivatives up to order k uniformly continuous in Ω. If Ω is omitted, $C^k = C^k(Q)$.

Let φ be a non-negative, real valued function in C_0^∞, vanishing outside $\{x \in \mathbb{R}^d : |x| < 1\}$, and such that $\int_{\mathbb{R}^d} \varphi(x)\,dx = 1$, and for $\varepsilon > 0$ define $\varphi_\varepsilon(x) = \varepsilon^{-d} \varphi(\varepsilon^{-1}x)$. The function

$$u^\varepsilon(x) = u * \varphi_\varepsilon(x) = \int_{\mathbb{R}^d} u(x - y)\,\varphi_\varepsilon(y)\,dy, \tag{3.1}$$

defined for functions u and points x whenever the right hand side makes sense, has many interesting well known properties, and we state some of them without proof, referring the reader to, e.g., the book by Ziemer [22, Theorem 1.6.1 and Remark 1.6.2].

Theorem 3.1 1. *If $u \in L^1_{loc}(\mathbb{R}^d)$, then for every $\varepsilon > 0$, $u^\varepsilon \in C^\infty(\mathbb{R}^d)$ and $\partial^\alpha u^\varepsilon = (\partial^\alpha \varphi_\varepsilon) * u$, for every multi-index α, where for $\alpha = (\alpha_1, \dots, \alpha_d)$, $\partial^\alpha = \partial_1^{\alpha_1} \dots \partial_d^{\alpha_d}$.*
2. *If $u \in L^1(\Omega)$ then $u^\varepsilon(x)$ is defined for $x \in \Omega$ and $\mathrm{dist}(x, \partial\Omega) > \varepsilon$.*
3. *$\lim_{\varepsilon \to 0} u^\varepsilon(x) = u(x)$ whenever x is a Lebesgue point of u.*

4. *If $u \in C(\Omega)$, and F is a compact subset of Ω, then u^ε converges uniformly to u on F as $\varepsilon \to 0$.*

The functions u^ε, also called regularizations or mollifiers of u, will make us work often with the sets

$$A_\varepsilon = \{x \in Q : \text{dist}(x, \partial Q) > \varepsilon\},$$

where $\varepsilon \in \mathbb{R}, 0 < \varepsilon < 1$.

Our first result is a simple characterization of continuous convex functions using the regularizations u^ε, and we omit its proof.

Lemma 3.2 *Suppose $u \in C$ and u^ε is defined as in* (3.1). *We have,*

1. *If u is convex, then u^ε is convex in A_ε for all $\varepsilon > 0$.*
2. *Conversely, if for a sequence of ε's converging to 0, u^ε is convex in A_ε, then u is convex.*

The Hessian of $u \in C^2$ at x is the symmetric matrix $Hu(x)$ whose ij entry is $\partial^2_{ij} u(x)$, and convex functions in C^2 are characterized by the positive semidefinite condition

$$Hu(x) \succeq 0 \quad \text{for all } x \in \mathring{Q}.$$

Also, for $u \in C^2$, and $\delta > 0$, the average

$$\overline{H}_\delta u(x) = \frac{1}{(2\delta)^d} \int\limits_{Q_\delta(x)} Hu(y) \, dy, \tag{3.2}$$

defined for $x \in A_\delta$, converges to $Hu(x)$ for all $x \in \mathring{Q}$, as $\delta \downarrow 0$. Thus the condition

$$\overline{H}_\delta u(x) \succeq 0 \quad \text{for all } x \text{ and } \delta \text{ such that } x \in A_\delta, \tag{3.3}$$

implies the convexity of u. Actually, since the average of positive semidefinite matrices is positive semidefinite, (3.3) is equivalent to convexity for $u \in C^2$.

The definitions of Hu and $\overline{H}_\delta u$ involve second order derivatives which are not defined if u is not smooth enough. However, we may express $\overline{H}_\delta u$ merely in terms of u and ∇u integrating by parts, and the resulting formula will make sense for $u \in C^1$. Thus, for $u \in C^1$ and $x \in A_\delta$ let us define $\widetilde{H}_\delta u(x)$ as the symmetric $d \times d$ matrix whose diagonal entries are

$$\left(\widetilde{H}_\delta u(x)\right)_{ii} = (2\delta)^{-d} \left(\int\limits_{F^+_{\delta,i}(x)} \partial_i u(y) \, dy - \int\limits_{F^-_{\delta,i}(x)} \partial_i u(y) \, dy \right), \tag{3.4a}$$

and, if $i \neq j$,

$$
\left(\widetilde{H}_\delta u(x)\right)_{ij} = (2\delta)^{-d} \left(\int_{F^+_{\delta,i}(x) \cap F^+_{\delta,j}(x)} u(y)\,\mathrm{d}y - \int_{F^+_{\delta,i}(x) \cap F^-_{\delta,j}(x)} u(y)\,\mathrm{d}y \right.
$$

$$
\left. - \int_{F^-_{\delta,i}(x) \cap F^+_{\delta,j}(x)} u(y)\,\mathrm{d}y + \int_{F^-_{\delta,i}(x) \cap F^-_{\delta,j}(x)} u(y)\,\mathrm{d}y \right), \quad (3.4\text{b})
$$

where $F^\pm_{\delta,i}(x)$ denotes the $d-1$ dimensional face of the cube $Q_\delta(x)$ having outward normal $\pm\mathbf{e}_i$.

Of course, since the Eq. (3.4) were obtained integrating by parts, we have:

Lemma 3.3 *If $u \in C^2$ then $\overline{H}_\delta u = \widetilde{H}_\delta u$.*

As we show now, the extension $\widetilde{H}_\delta$ of $\overline{H}_\delta$ to functions in C^1, still gives a local characterization of convexity.

Theorem 3.4 *The following are valid if $u \in C^1$:*

1. *If u is convex, then $\widetilde{H}_\delta u(x) \succeq 0$ for all $\delta > 0$ and $x \in A_\delta$.*
2. *If for a sequence $\delta_n \downarrow 0$, $\widetilde{H}_{\delta_n} u(x) \succeq 0$ for all $x \in A_{\delta_n}$, then u is convex.*

Proof Suppose $u \in C^1$, and consider the regularization u^ε defined in (3.1). Since we can interchange derivatives and integrals with the convolution,

$$
\widetilde{H}_\delta u^\varepsilon(x) = \widetilde{H}_\delta(u * \varphi_\varepsilon)(x) = \left(\widetilde{H}_\delta u\right) * \varphi_\varepsilon(x) \quad \text{for } x \in A_{\delta+\varepsilon}. \tag{3.5}
$$

If u is convex, by the first part of Lemma 3.2 we know that u^ε is convex in A_ε, and since $u^\varepsilon \in C^\infty$, for $x \in A_{\delta+\varepsilon}$ we have $\overline{H}_\delta u^\varepsilon(x) \succeq 0$. Thus, by Lemma 3.3,

$$
\widetilde{H}_\delta u^\varepsilon(x) \succeq 0 \quad \text{for } x \in A_{\delta+\varepsilon}.
$$

Using Theorem 3.1, since u^ε and their derivatives converge uniformly to u in compact subsets of $\mathring{Q}$, we must have $\widetilde{H}_\delta u(x) \succeq 0$ for x in compact subsets of A_δ, and therefore in the whole of A_δ.

On the other hand, if $\widetilde{H}_\delta u(x) \succeq 0$ for $x \in A_\delta$, since an integral mean of positive semidefinite matrices is positive semidefinite, using (3.5) and Lemma 3.3, we see that

$$
\overline{H}_\delta u^\varepsilon(x) = \widetilde{H}_\delta u^\varepsilon(x) \succeq 0 \quad \text{for } x \in A_{\varepsilon+\delta},
$$

which implies that u^ε is convex in $A_{\varepsilon+\delta}$. Letting δ go to 0 while keeping ε fixed, we see that u^ε is convex in A_ε, and by the second part of Lemma 3.2, u must be convex. □

4 Approximating convex functions, the discrete Hessian

There are two main issues when defining the set of discrete approximants to be used:

1. we want it to be rich enough to approximate every convex function, and
2. we want this set to be not too large, to avoid convergence to non-convex functions.

The first point is very natural, and necessary to approximate the solution to the problem. The second point might look artificial at first sight, but if not enforced, then we could be approximating a non-convex function. This is the case, for instance, if we only require convexity along the coordinate axes, as we stated in Sect. 2.

Our discrete approximations will be a subset of functions having positive semidefinite discrete Hessians, and the main purpose of this section is to address the two issues mentioned above.

We will show in Theorem 4.2 that, despite the examples 2.3 and 2.4, the discrete Hessian H_h may be used to obtain very good approximations to convex functions of continuous variables.

This is not surprising since we can approximate, say, $u \in C^3$, by discrete functions whose finite differences up to order 2 converge to the derivatives up to order 2 of u, and then add a small perturbation to bring up the eigenvalues of the discrete Hessians. This is the main idea of the proof, and in its course, we will use the following part of the Hoffman–Wielandt Theorem [10].

Theorem 4.1 (Hoffman and Wielandt) *There exists a positive constant c_d, depending only on the dimension d, such that if $A = [a_{ij}]$ and $B = [b_{ij}]$ are symmetric $d \times d$ matrices, and λ and μ are their minimum eigenvalues, then*

$$|\lambda - \mu| \leq c_d \max_{ij} |a_{ij} - b_{ij}|.$$

Theorem 4.2 *Let $u \in C^3$ be convex. Then, for any $\varepsilon > 0$, there exists $h_0 = h_0(u, \varepsilon) > 0$ such that for any h, $0 < h < h_0$, there exists a function u_h defined on $\mathcal{M}_h$ satisfying for all $x \in \mathcal{M}_h$,*

$$|u_h(x) - u(x)| + \sum_{i=1}^{d} |\Delta_{h,i} u_h(x) - \partial_i u(x)|$$
$$+ \sum_{1 \leq i \leq j \leq d} |\Delta_{h,ij}^2 u_h(x) - \partial_{ij}^2 u(x)| < \varepsilon, \qquad (4.1)$$

and

$$H_h u_h(x) \succeq 0 \quad \text{for all } x \in \mathring{\mathcal{M}}_h.$$

Let us recall that in the inequality (4.1), for $x \in \partial \mathcal{M}_h$ we consider only the finite differences that are defined at that x.

Proof If u is convex and smooth, given $\varepsilon_1 > 0$ there exists $\delta_1 > 0$ such that for all $x \in Q$ and $0 < h < \delta_1$,

$$\sum_{i=1}^{d} |\Delta_{h,i}\, u(x) - \partial_i u(x)| + \sum_{1 \le i \le j \le d} |\Delta_{h,ij}^2\, u(x) - \partial_{ij}^2 u(x)| < \frac{\varepsilon_1}{c_d},$$

where c_d is the constant in Theorem 4.1. Hence, since u is convex and therefore $Hu(x) \succeq 0$, the minimum eigenvalue of $H_h u(x)$ is uniformly bounded below by $-\varepsilon_1$.

Now consider the function

$$g(x) = \frac{1}{2}\,|x|^2,$$

for which, if h small enough,

$$|g(x)| \le d/2, \quad |\Delta_{h,i}\, g(x)| = |x_i + h/2| \le 2, \quad \text{and} \quad H_h g(x) = Hg(x) = I_d,$$

where I_d is the identity matrix in $\mathbb{R}^{d \times d}$.

If $0 < h < \min\{\varepsilon_1, \delta_1\}$, the function

$$u_h = u + \varepsilon_1 g,$$

defined on $\mathcal{M}_h$, satisfies

$$H_h u_h(x) \succeq 0 \quad \text{for } x \in \overset{\circ}{\mathcal{M}}_h,$$
$$|u_h(x) - u(x)| \le \frac{d\varepsilon_1}{2},$$

and

$$\sum_{i=1}^{d} |\Delta_{h,i}\, u_h(x) - \partial_i u(x)| + \sum_{1 \le i \le j \le d} |\Delta_{h,ij}^2\, u_h(x) - \partial_{ij}^2 u(x)|$$
$$\le \sum_{i=1}^{d} |\Delta_{h,i}(u_h(x) - u(x))| + \sum_{i=1}^{d} |\Delta_{h,i}\, u(x) - \partial_i u(x)|$$
$$+ \sum_{1 \le i \le j \le d} |\Delta_{h,ij}^2(u_h(x) - u(x))| + \sum_{1 \le i \le j \le d} |\Delta_{h,ij}^2\, u(x) - \partial_{ij}^2 u(x)|$$
$$\le 2d\varepsilon_1 + d\varepsilon_1 + \frac{\varepsilon_1}{c_d}.$$

Thus, for some constant c_d' depending only on d, the inequality (4.1) holds with ε replaced by $c_d'\, \varepsilon_1$.

The result follows now by taking ε_1 and h_0 appropriately. $\qquad \square$

The main implication of Theorem 4.2 is that any smooth convex function is a limit of a sequence of functions with positive semidefinite discrete Hessians, giving an affirmative answer to the first issue mentioned at the beginning of this section. Moreover, an application of Lemma 3.2 implies this result also for non-smooth convex functions, with convergence in $\| \cdot \|_0$ on compact subsets of Ω (see definitions below).

We would like to show next that our set of approximants also solves the second issue. That is, if we have a convergent (in certain norm) sequence of functions with positive semidefinite discrete Hessians, then the limit is convex. On the other hand, if the sequence is not convergent, we would like also to understand under which further conditions on the sequence we may extract a subsequence converging to a convex function.

In what follows, we will work with sequences of functions defined on finer and finer meshes, that is, sequences $S = (u_{h_n})_{n \in \mathbb{N}}$ with $u_{h_n} \in \mathcal{U}_{h_n}$ for every n, such that $h_n / h_{n+1} \in \mathbb{N}$ and $(h_n)_n$ decreases to 0. We will denote by $\mathcal{S}$ the set of all such sequences, and use the notation

$$\mathcal{M}(S) = \cup_n \mathcal{M}_{h_n}, \quad \mathring{\mathcal{M}}(S) = \mathcal{M}(S) \cap \mathring{Q}, \quad \text{and} \quad \partial \mathcal{M}(S) = \mathcal{M}(S) \cap \partial Q.$$

So as not to clutter even more the notation, usually we will drop the index n, writing $S = (u_h)$ for $S \in \mathcal{S}$, when this does not lead to confusion.

If $S = (u_h) \in \mathcal{S}$ and $u : Q \to \mathbb{R}$, we will say that

$$\lim_{h \to 0} u_h(x) = u(x) \quad \text{uniformly for } x \in \mathcal{M}(S),$$

if for any $\varepsilon > 0$, there exists $h_0 = h_0(\varepsilon) > 0$, so that $|u_h(x) - u(x)| < \varepsilon$ for all $h < h_0$ and $x \in \mathcal{M}_h$, that is, $\max_{x \in \mathcal{M}_h} |u_h(x) - u(x)| \to 0$ as $h \to 0$. In this case we write, with a little abuse of notation,

$$\lim_{h \to 0} \|u_h - u\|_0^* = 0.$$

If in addition $u \in C^1$, we write, similarly,

$$\lim_{h \to 0} \|u_h - u\|_1^* = 0,$$

to indicate that $\lim_{h \to 0} \|u_h - u\|_0^* = 0$ and

$$\lim_{h \to 0} \Delta_{h,i} u_h(x) = \partial_i u(x)$$

uniformly for all $x \in \mathcal{M}(S)$ for which the finite differences make sense and $i = 1, \ldots, d$.

Lemma 4.3 *Suppose $S = (u_h) \in \mathcal{S}$ and $u \in C^1$ are such that*

$$\lim_{h \to 0} \|u_h - u\|_1^* = 0,$$

and

$$H_h u_h(x) \succeq 0 \quad \text{for all } x \in \mathring{\mathcal{M}}(S).$$

Then u is convex.

Proof We follow closely what was done in Sect. 3, using a variant of the divergence theorem for discrete variables, so that only approximations to the function or its derivatives—but not the second derivatives—are needed.

For $x \in \mathring{\mathcal{M}}_{h'}$ and $h \ll h'$, let us define

$$H^*_{h',h} u_h(x) = \sum_y H_h u_h(y),$$

where the sum is over all $y \in \mathcal{M}_h$ such that $Q_h(y) \subset Q_{h'}(x)$.

Since the sum of positive semidefinite matrices is a positive semidefinite matrix, we have

$$H^*_{h',h} u_h(x) \succeq 0 \quad \text{for all } x \in \mathring{\mathcal{M}}_{h'}.$$

In one dimension we have just one entry in $H^*_{h',h}$, involving a term of the form

$$
\begin{aligned}
&u(x - h') - 2\,u(x - h' + h) + u(x - h' + 2h) \\
&\quad + u(x - h' + h) - 2\,u(x - h' + 2h) + u(x - h' + 3h) + \cdots \\
&\quad + u(x + h' - 2h) - 2\,u(x + h' - h) + u(x + h') \\
&= \bigl(u(x + h') - u(x + h' - h)\bigr) - \bigl(u(x - h' + h) - u(x - h')\bigr).
\end{aligned}
$$

That is, if $d = 1$ then

$$H^*_{h',h} u_h(x) = \left[\tfrac{1}{h} \left(\Delta_h u_h(x + h' - h) - \Delta_h u_h(x - h') \right) \right]. \tag{4.2}$$

If $d = 2$ and $x = (x_1, x_2)$, the diagonal entries are similar to the one dimensional case. For instance,

$$
\left(H^*_{h',h} u_h(x) \right)_{11} = \frac{1}{h} \left(\sum_k \Delta_{h,1} u_h(x_1 + h' - h, x_2 + kh) \right.
$$
$$
\left. - \sum_k \Delta_{h,1} u_h(x_1 - h', x_2 + kh) \right), \tag{4.3}
$$

where both sums are on k's such that $-h' < kh < h'$. On the other hand, the off-diagonal terms are of the form

$$
\begin{aligned}
\left(H^*_{h',h} u_h(x)\right)_{12} = \frac{1}{4h^2} & \left((u_h(x + \alpha) + u_h(x + \beta) + u_h(x + \gamma) + u_h(x + \delta))\right. \\
& - \left(u_h(x + \alpha') + u_h(x + \beta') + u_h(x + \gamma') + u_h(x + \delta')\right) \\
& - \left(u_h(x - \alpha') + u_h(x - \beta') + u_h(x - \gamma') + u_h(x - \delta')\right) \\
& + \left.(u_h(x - \alpha) + u_h(x - \beta) + u_h(x - \gamma) + u_h(x - \delta))\right),
\end{aligned}
$$

(4.4)

where

$$
\begin{aligned}
\alpha &= (h' - h)\,\mathbf{e}_1 + (h' - h)\,\mathbf{e}_2, & \alpha' &= (h' - h)\,\mathbf{e}_1 - (h' - h)\,\mathbf{e}_2, \\
\beta &= h'\,\mathbf{e}_1 + (h' - h)\,\mathbf{e}_2, & \beta' &= h'\,\mathbf{e}_1 - (h - h')\,\mathbf{e}_2, \\
\gamma &= (h' - h)\,\mathbf{e}_1 + h'\,\mathbf{e}_2, & \gamma' &= (h' - h)\,\mathbf{e}_1 - h'\,\mathbf{e}_2, \\
\delta &= h'\,\mathbf{e}_1 + h'\,\mathbf{e}_2, & \delta' &= h'\,\mathbf{e}_1 - h'\,\mathbf{e}_2.
\end{aligned}
$$

For $d > 2$ we get similar expressions to those of Eqs. (4.2) and (4.3) for the diagonal terms, and (4.4) for the off-diagonal terms, except that—as when going from (4.2) to (4.3)—they must be summed over mesh points on $d - 1$ dimensional surfaces perpendicular to the (say) ith direction, or $d - 2$ dimensional surfaces perpendicular to the (say) ij plane.

This was also the case in the Eq. (3.4) for the continuous case, and so we can think of these sums as approximations to those integrals, where small d-dimensional cubes of side length h centered at mesh points have been used. In order to do this we must multiply the entries by h^d so as to obtain the correct dimensions.

In fact, since we are assuming

$$
\lim_{h \to 0} \|u_h - u\|_1^* = 0,
$$

and $H^*_{h',h} u_h(x)$ involves either no finite differences (off the diagonal) or only first order differences (in the diagonal) which converge uniformly to (respectively) u or its first derivatives,

$$
\lim_{h \to 0} h^d\, H^*_{h',h} u_h(x) = R_{h'} u(x) \quad \text{for all } x \in \mathring{\mathcal{M}}_{h'},
$$

where

$$
R_{h'} u(x) \succeq 0,
$$

since the limit of positive semidefinite matrices is positive semidefinite.

On the other hand, due to the convergence of u_h to u in the $\|\cdot\|_1^*$ norm, we can see that, as $h \downarrow 0$,

$$\left(h^d \, H_{h',h}^* u_h(x)\right)_{ii} \to \int_{F_{\delta,i}^+(x)} \partial_i u(y) \, dy - \int_{F_{\delta,i}^-(x)} \partial_i u(y) \, dy,$$

$$\left(h^d \, H_{h',h}^* u_h(x)\right)_{ij} \to \left(\int_{F_{\delta,i}^+(x) \cap F_{\delta,j}^+(x)} u(y) \, dy - \int_{F_{\delta,i}^+(x) \cap F_{\delta,j}^-(x)} u(y) \, dy \right.$$

$$\left. - \int_{F_{\delta,i}^-(x) \cap F_{\delta,j}^+(x)} u(y) \, dy + \int_{F_{\delta,i}^-(x) \cap F_{\delta,j}^-(x)} u(y) \, dy \right),$$

if $i \neq j$. Thus,

$$R_{h'} u(x) = (2h')^d \, \widetilde{H}_{h'} u(x) \quad \text{for } x \in \mathring{\mathcal{M}}_{h'},$$

which implies $\widetilde{H}_{h'} u(x) \succeq 0$ for all $x \in \mathring{\mathcal{M}}_{h'}$. The result now follows from Theorem 3.4. $\qquad\square$

The convergence in $\|\cdot\|_1^*$ should be guaranteed by the definition of the discrete problem and is problem dependent. We make now a general assumption on an algorithm to ensure convergence in $\|\cdot\|_1^*$.

Let us denote by Λ^1 the space of Lipschitz continuous functions defined on Q, recalling that $\Lambda^1 = W^{1,\infty}$, the space of continuous functions having derivatives in the weak sense up to order one bounded. If $K > 0$, we let Λ_K^1 be the set of functions $u \in \Lambda^1$ such that both $|u(x)| \leq K$ for all $x \in Q$ and $|u(x) - u(x + t\,\mathbf{e}_i)| \leq Kt$ for all $i = 1, \ldots, d$ and $t > 0$ whenever $x, x + t\,\mathbf{e}_i \in Q$. Of course, the condition $|u(x) - u(x + t\,\mathbf{e}_i)| \leq Kt$ is equivalent to $|\partial_i u(x)| \leq K$.

Similarly, let us denote by $\Lambda^2 = W^{2,\infty}$ the space of functions whose first derivatives are in Λ^1, and by Λ_K^2 the set of all functions in Λ^2 which have all weak derivatives up to order 2 bounded by K. In particular, we have $\Lambda^2 \subset C^1$.

By analogy to the continuous case, let us define for h and K positive the following spaces of discrete functions:

$$\begin{aligned} \Lambda_{h,K}^0 &= \{u \in \mathcal{U}_h : |u(x)| \leq K, x \in \mathcal{M}_h\}, \\ \Lambda_{h,K}^1 &= \{u \in \Lambda_{h,K}^0 : |\Delta_{h,i} \, u(x)| \leq K, x \in \mathcal{M}_h, i = 1, \ldots, d\}, \qquad (4.5) \\ \Lambda_{h,K}^2 &= \{u \in \Lambda_{h,K}^1 : |\Delta_{h,ij}^2 \, u(x)| \leq K, x \in \mathcal{M}_h, i, j = 1, \ldots, d\}, \end{aligned}$$

with the understanding that on the right hand sides we take $\Delta_{h,i} \, u(x)$ or $\Delta_{h,ij}^2$ for all x, i and j where it makes sense.

The following result is a version of the Arzela–Ascoli theorem, and we omit its proof.

Theorem 4.4 *If $S = (u_h) \in \mathcal{S}$ for which $u_h \in \Lambda_{h,K}^1$ for all $h > 0$, then there exists $u \in \Lambda_K^1$ and a subsequence $S' = (u_{h'})$ of S such that*

$$\lim_{h' \to 0} \|u_{h'} - u\|_0^* = 0.$$

Combining the previous results, we have:

Theorem 4.5 *If $S = (u_h) \in \mathcal{S}$ is such that*

$$u_h \in \Lambda_{h,K}^2 \quad and \quad H_h u_h(x) \succeq 0$$

for all $x \in \overset{\circ}{\mathcal{M}}(S)$ and $h > 0$, then there exists $u \in \Lambda_K^2$, and a subsequence $S' = (u_{h'})$ of S such that

$$\lim_{h' \to 0} \|u_{h'} - u\|_1^* = 0 \quad and \quad u \text{ is convex.}$$

Proof Applying Theorem 4.4 to the functions $\Delta_{h,i}$, perhaps on some smaller meshes $\mathcal{M}_{i,h}'$ with $\overset{\circ}{\mathcal{M}}_h \subset \mathcal{M}_{i,h}' \subset \mathcal{M}_h$, we may find functions $u_i \in \Lambda_K^1$ and a subsequence of S, $S' = (u_{h'})$, such that

$$\lim_{h' \to 0} \|\Delta_{h',i}\, u_{h'} - u_i\|_0^* = 0 \quad \text{for } i = 1, \ldots, d.$$

To show that there exists u such that $u_i = \partial_i u$ (in the classical sense), we define for $x = (x_1, \ldots, x_d) \in Q$,

$$u\left(\sum_{i=1}^d x_i\, \mathbf{e}_i \right) = u(0) + \sum_{j=1}^d \int_0^{x_j} u_j\left(t_j\, \mathbf{e}_j + \sum_{i=1}^{j-1} x_i\, \mathbf{e}_i \right) dt_j.$$

Since for $i = 1, \ldots, d$, u_i is bounded (by K) and continuous, u is well defined and continuous. Using that $\Delta_{h',i}\, u_{h'}$ converge uniformly to u_i, for $x \in \mathcal{M}(S')$ we may write

$$u(x) = \lim_{h' \to 0} u_{h'}(0) + \sum_{j=1}^d \sum_{k_j=0}^{x_j/h'} h'\, \Delta_{h',j}\, u_{h'}\left(k_j h'\, \mathbf{e}_j + \sum_{i=1}^{j-1} x_i\, \mathbf{e}_i \right).$$

But, for a given h' and all $x \in \mathcal{M}_{h'}$ we have

$$u_{h'}(0) + \sum_{j=1}^d \sum_{k_j=0}^{x_j/h'} h'\, \Delta_{h',j}\, u_{h'}\left(k_j h'\, \mathbf{e}_j + \sum_{i=1}^{j-1} x_i\, \mathbf{e}_i \right)$$

$$= u_{h'}(0) + \sum_{j=1}^d \left(u_{h'}\left(\sum_{i=1}^j x_i\, \mathbf{e}_i \right) - u_{h'}\left(\sum_{i=1}^{j-1} x_i\, \mathbf{e}_i \right) \right) = u_{h'}(x), \quad (4.6)$$

and therefore

$$\lim_{h' \to 0} \|u_{h'} - u\|_0^* = 0.$$

Arguing as in Eq. (4.6) and taking limits, we may also verify the "independence of path", that is, for $x \in \mathcal{M}_{h_0'}$ and $h' \le h_0'$ we may write,

$$u(x + h' \, \mathbf{e}_i) - u(x) = \int_0^{h'} u_i(x + t_i \, \mathbf{e}_i) \, dt_i \quad \text{for all } i = 1, \dots, d.$$

Using the continuity of u, we see that the last equation is valid for all $x \in \overset{\circ}{Q}$ and h' small enough, and, using once more the continuity of u, that

$$u(x + \delta \, \mathbf{e}_i) - u(x) = \int_0^{\delta} u_i(x + t_i \, \mathbf{e}_i) \, dt_i \quad \text{for all } i = 1, \dots, d,$$

for all $x \in \overset{\circ}{Q}$ and all $\delta > 0$, and therefore

$$\partial_i u(x) = u_i(x) \quad \text{for all} \quad x \in \overset{\circ}{Q} \quad \text{and} \quad i = 1, \dots, d.$$

The convexity of u follows from Lemma 4.3. $\qquad \square$

Remark 4.6 If u_n is a sequence of convex functions converging pointwise to u, then u must be convex. However it is not true in general that the second derivatives or the Hessian of u_n will converge to that of u, even if we have uniform convergence of u_n and its derivatives to those of u. In this sense, Theorem 4.5 cannot be bettered too much.

For example, consider in $d = 1$ the functions

$$u(x) = \frac{x^2}{2},$$

and, with $h = 1/n$ for $n \in \mathbb{N}$,

$$u_h(x) = u(x) + \frac{\cos(\pi n x)}{(n\pi)^2}.$$

We have,

$$u_h(x) \ge 0, \quad \lim_{h \to 0} u_h(x) = u(x),$$
$$0 \le u_h'(x) = u'(x) - \frac{\sin(\pi n x)}{n\pi} \le 1, \quad \lim_{h \to 0} u_h'(x) = u'(x),$$

where the limits are uniform in x. Also,

$$u_h''(x) = u''(x) - \cos(\pi n x) \ (\geq 0), \qquad \text{and} \qquad \Delta^2 u_h(x) = 1 - \frac{4\cos(\pi n x)}{\pi^2}.$$

Taking $n = 2^{j+m}$, $x = \frac{i}{2^j} = \frac{2^m i}{n}$, for $m > 0$ we obtain

$$\Delta^2 u_h(x) = 1 - \frac{4\cos(\pi 2^m i)}{\pi^2} = 1 - \frac{4}{\pi^2} \approx 0.594715,$$

and so the second order differences at dyadic points converge to this constant value. However, also at these points, $u''(x) = 1$. $\qquad\square$

On the other hand, we may weaken the conditions on convergence, following what was done for the one-dimensional case in Sect. 2.

If u_h is defined in $\mathcal{M}_h$ and $H_h u_h(x) \succeq 0$ for $x \in \mathcal{M}_h$, its restriction to the one-dimensional line obtained by fixing all coordinates except the i-th one, satisfies

$$\frac{u(x + ih\,\mathbf{e}_i) - u(x + (i-1)\,h\,\mathbf{e}_i)}{h} \leq \frac{u(x + (i+1)\,h\,\mathbf{e}_i) - u(x + ih\,\mathbf{e}_i)}{h},$$

for $i = 1, \ldots, d$, since these are coefficients in the main diagonal of $H_h u_h(x + h\mathbf{e}_i)$. That is, the inequalities (2.2) are satisfied, and hence if for some $M > 0$,

$$|u_h(x)| \leq M \quad \text{for all } x \in \mathcal{M}_h,$$

then for any given $\varepsilon > 0$ we may find a constant C, depending on M and ε but not on u, such that

$$|\Delta_{h,i}\, u(x)| \leq C \quad \text{for all } x \in \mathcal{M}_h \text{ such that } \mathrm{dist}(x, \partial Q) > \varepsilon.$$

Therefore, by applying Theorem 4.4 to d-dimensional cubes contained in $\mathring{Q}$, we may strengthen Lemma 4.3 to obtain a variant of Lemma 2.1:

Corollary 4.7 *Let $S = (u_h) \in \mathcal{S}$ such that*

$$H_h u_h(x) \succeq 0 \ \ \text{for all } x \in \mathring{\mathcal{M}}(S),$$

and suppose u is a (finite) function defined on $\mathcal{M}(S)$ satisfying

$$\lim_{h\downarrow 0} u_h(x) = u(x) \ \ \text{for all } x \in \mathcal{M}(S).$$

Then, the convergence of u_h to u is uniform on compact subsets of $\mathring{Q}$, and u may be uniquely extended to a convex function defined on all of Q.

Similarly, we may weaken the conditions of Theorem 4.4 to obtain the following version of Lemma 2.2:

Corollary 4.8 *If $M > 0$ and $S = (u_h) \in \mathcal{S}$ are such that*

$$|u_h(x)| \leq M \quad \text{for all } h \text{ and } x \in \mathcal{M}_h.$$

Then there exists a continuous convex function u defined on Q and a subsequence $S' = (u_{h'})$ of S such that

$$\lim_{h' \to 0} u_{h'}(x) = u_h(x) \quad \text{for all } x \in \mathcal{M}(S').$$

Moreover, the convergence is uniform on compact subsets of $\mathring{Q}$.

We conclude this section with some comments.

Remark 4.9 (boundary behavior). If $u \in C^2$, $\partial_{ij}^2 u(x)$ is defined initially for $x \in \mathring{Q}$, but the very definition of C^2 as the set of those functions in $C^2(\mathring{Q})$ having second order derivatives uniformly continuous on $\mathring{Q}$, makes it possible to define $\partial_{ij}^2 u(x)$ for $x \in \partial Q$ by a limiting argument. And the same goes for lower order derivatives, and even $Hu(x)$.

The situation is different for discrete functions defined on $\mathcal{M}_h$ for fixed $h > 0$, since we cannot take limits. However, the particular geometry of Q makes it possible (as we did) to consider for $x \in \partial \mathcal{M}_h$ as many finite differences as we can. For instance, $\Delta_{ii}^2 u(x)$ is well defined for $x \in \partial \mathcal{M}_h$ as long as $x \pm h\, \mathbf{e}_i \in \mathcal{M}_h$.

Pushing the definitions a little further, we may define the discrete Hessian $H_h u(x)$ for $x \in \partial \mathcal{M}_h$, by including as many second derivatives as we can. For example, if $d = 3$ and $h = 1/2$, we can define

$$H_h u(x) = \begin{bmatrix} \Delta_{h,11}^2 u(x) & \Delta_{h,12}^2 u(x) \\ \Delta_{h,12}^2 u(x) & \Delta_{h,22}^2 u(x) \end{bmatrix} \qquad \text{if } x = (1/2, 1/2, 0),$$

$$H_h u(x) = \begin{bmatrix} \Delta_{h,11}^2 u(x) \end{bmatrix} \qquad\qquad\qquad \text{if } x = (1/2, 0, 0),$$

leaving $H_h u(x)$ undefined if $x = (0, 0, 0)$.

As the reader may verify, our previous results involving H_h remain valid with this interpretation of the discrete Hessian.

5 Approximating functionals

We are in position now to use finite difference approximations of a wide class of optimization problems on convex functions.

Let us describe this technique by assuming, for instance, that $(V, \|\cdot\|_V)$ is a Banach space of real valued functions on Q, the functional

$$J(v) = \int_Q F(x, v(x), \nabla v(x)) \, \mathrm{d}x$$

is defined and continuous on V, and we are interested in the optimization problem

$$\inf \{J(v) : v \in \mathcal{C}\}, \tag{5.1}$$

where $\mathcal{C}$ is a family of convex functions, $\mathcal{C} \subset V$.

If the functions in $\mathcal{C}$ may be approximated by convex functions in $C^3 \cap V$, then using Theorem 4.2 it may be not too difficult to define for each $h > 0$ (or a sequence converging to 0 of such h's), a family $\mathcal{C}_h$, $\mathcal{C}_h \subset \mathcal{U}_h$, and a functional J_h defined on $\mathcal{C}_h$, such that:

1. $H_h v_h(x) \succeq 0$ for all $v_h \in \mathcal{C}_h$ and $x \in \mathring{\mathcal{M}}_h$,
2. for any $v \in \mathcal{C}$ and any $\varepsilon > 0$, there exists $h > 0$ and $v_h \in \mathcal{C}_h$ such that $|J_h(v_h) - J(v)| < \varepsilon$.

Condition 2 immediately implies that

$$\inf \{J(v) : v \in \mathcal{C}\} \geq \inf_h \inf \{J_h(v_h) : v_h \in \mathcal{C}_h\}. \tag{5.2}$$

To prove the converse, we observe that

$$\inf_h \{J_h(v_h) : v_h \in \mathcal{C}_h\} = \inf_{h,K>0} \inf \{J_h(v_h) : v_h \in \mathcal{C}_{h,K}\},$$

where $\mathcal{C}_{h,K} = \mathcal{C}_h \cap \Lambda^2_{h,K}$. Keeping $K > 0$ fixed, we may find a sequence $S = (u_h^K)$, with $u_h^K \in \mathcal{C}_{h,K}$ such that

$$J_h(u_h^K) \downarrow \inf_h \inf \{J_h(v_h) : v_h \in \mathcal{C}_{h,K}\},$$

and using Theorem 4.5, a subsequence $S' = (u_{h'}^K)$ and a convex function $u^K \in C^2$ with $\|u_h^K - u^K\|_1^*$ converging to 0.

If V_h and J_h are such that

$$u^K \in \mathcal{C} \quad \text{and} \quad J_h(u_h^K) \to J(u^K),$$

letting $K \to \infty$ we will have

$$\inf \{J(v) : v \in \mathcal{C}\} \leq \inf_h \inf \{J_h(v_h) : v_h \in \mathcal{C}_h\}. \tag{5.3}$$

Putting together the inequalities (5.2) and (5.3), we will have discrete approximations of the problem (5.1).

Let us give some more concrete examples. Suppose, for instance that $V = H^1$ is the set of functions $u : Q \to \mathbb{R}$ with finite norm

$$\|u\| = \left(\int_Q |\nabla u(x)|^2 + |u(x)|^2 \, dx \right)^{1/2},$$

and suppose $\mathcal{C}$ is the set of all convex functions in H^1. Given $f \in V$ we would like to find its projection on $\mathcal{C}$, that is, find $u \in \mathcal{C}$ such that

$$\|u - f\| = \min_{v \in \mathcal{C}} \|v - f\|,$$

and, getting rid of the square roots, we may set

$$J(v) = \|v - f\|^2.$$

In this example we actually have a unique minimum, since the norm is strictly convex. We may consider then $\mathcal{C}_h$ as the set of discrete functions $v_h \in \mathcal{U}_h$ with $H_h v_h(x) \succeq 0$ for all $x \in \mathring{\mathcal{M}}_h$.

Assuming for simplicity that $f \in C^1$, for $v_h \in \mathcal{U}_h$ we define

$$J_h(v_h) = h^d \left(\sum_{x \in \mathring{\mathcal{M}}_h} \left(|v_h(x) - f(x)|^2 + \sum_{i=1}^{d} |\Delta_{h,i} \, v_h(x) - \partial_i f(x)|^2 \right) \right).$$

Then, it is easy to see that (5.2) and (5.3) hold. In fact, the convex functions of V may be approximated by C^3 convex functions, so that, given $\varepsilon > 0$ there exists $v \in C^3 \cap H^1$ such that

$$\|u - v\| < \frac{1}{2} \varepsilon,$$

and use Theorem 4.2 to find h small enough and $v_h \in \mathcal{C}_h$ such that

$$|J_h(v_h) - J(v)| < \frac{1}{2} \varepsilon.$$

Thus we will have

$$|J_h(v_h) - J(u)| \leq |J_h(v_h) - J(v)| + |J(v) - J(u)| < \varepsilon,$$

for some $v_h \in \mathcal{C}_h$.

6 Numerical results

In this section we illustrate the behavior of the numerical scheme by applying it to the problems mentioned in the introduction, namely, the monopolist problem, norm projections on the set of convex functions, and fitness of data by discrete convex functions.

In all of the following examples, we associate with each mesh node P_k, $k = 1, \ldots, N$, the unknown value u_k, and the square $Q_k = Q \cap Q_h(P_k)$, of area $|Q_k|$. We also consider the discrete Hessian $H_h(P_k)$ as discussed in the Remark 4.9, i.e., imposing convex constraints on the boundary whenever they make sense.

Even though Theorem 4.5 requires us to impose upper bounds on the second order differences in order to ensure convergence, the following numerical experiments were carried on without this requirement. Convergence was nevertheless observed at optimal rates.

The times reported correspond to the experiments being run on a PC, with a 2.8 GHz Pentium IV processor and 2 GB of RAM, running *Linux*. The matrices were assembled using *OCTAVE* [7] and the semidefinite program was solved using *CSDP* 5 [2] with the default parameters. The graphics were obtained using *Mathematica* [21].

6.1 The monopolist problem

Since this problem is not widely known in the mathematics community, let us start by giving a brief description of it following the one given in [14], where it is referred to as the *revenue maximization in a multiple-good monopoly* problem.

Problem 6.1 (The monopolist problem) A seller with d different objects faces a single buyer, whose preferences over consumption and money transfers are given by $U(x, p, t) = x \cdot p - t$, where $x \in [0, 1]^d$ is the vector of the buyer's valuations, $p \in \mathbb{R}^d$ is the vector of quantities consumed for each good, and $t \in \mathbb{R}$ is the monetary transfer made to the seller. The valuations x are only observed by the buyer, and a density function $f(x)$ represents the seller's belief on the buyer's private information x. The seller's problem is to design a revenue maximizing mechanism to carry out the sale, and it is enough to consider only direct revelation mechanisms: the buyer must prefer to reveal its information truthfully (incentive compatibility) and to participate voluntarily (individual rationality). Under these conditions, it can be proved that the seller's problem may be written as

$$\max_{u \in \mathcal{C}} \int_Q (\nabla u(x) \cdot x - u(x)) \ f(x) \, dx,$$

where $Q = [0, 1]^d$, f is a non-negative probability density function over Q, and $\mathcal{C}$ is the set of functions u satisfying

1. u is convex,
2. $0 \leq \nabla u(x) \leq 1$ for all $x \in Q$ (the gradient taken in the weak sense and the inequalities componentwise), and
3. $u(0) = 0$.

The functional to be maximized is the seller's expected revenue. For a buyer of type x, the solution u to this optimization problem represents the utility received by her, and the i-th component of ∇u denotes the probability that she will obtain good i. The restriction of convexity stems from incentive compatibility, and the condition $u \geq 0$ from individual rationality.

We show now some numerical results for the problem 6.1 for the special case $f = 1$, for which analytic solutions in 2 and 3 dimensions are known, allowing us to judge

the behavior of the discrete approximations. That is, the functional to be minimized is

$$J(u) = \int_Q (u(x) - \nabla u(x) \cdot x) \, dx.$$

In both 2 and 3 dimensions, the solutions are piecewise linear convex functions whose partial derivatives are either 0 or 1. For example, for $d = 2$ the solution is

$$u(x_1, x_2) = \max \{0, x_1 - a, x_2 - a, x_1 + x_2 - b\},$$

where

$$a = \frac{2}{3} \quad \text{and} \quad b = \frac{1}{3} \left(4 - \sqrt{2}\right),$$

and the value at the optimum is

$$J(u) = \frac{2}{27} \left(6 + \sqrt{2}\right) \approx 0.549201.$$

In Fig. 1 we show the contour lines obtained. In (a) the analytic solution, and then the discrete solution for $h = 1/16$, $1/32$ and $1/64$ in, respectively, (b), (c) and (d), with the contours of the analytic solution shown in a lighter gray. The contours are $10^{-7}, 0.1, 0.2, \ldots, 1.1$ (the CSDP solution is always positive due to the way it is solved).

We observe that the scheme introduces quite a bit of diffusion at lower resolutions, and, in particular, the heights at the point $(1, 1)$ increase to the analytic solution at that point as h decreases. However, the jump in the gradients is well captured, especially at higher resolutions, even though we are asking for conditions on the discrete Hessians which are unbounded from above.

In Table 1 we give some comparative results at different mesh sizes. In this table we indicate by $J_h(I_h u)$ the discrete functional evaluated at $I_h(u)$, the interpolant in $\mathcal{M}_h$ of the exact solution, and the error column refers to $\max |u_h - I_h u|$ (L^∞ error).

It is interesting to notice that even though the solution is not smooth (u is only Lipschitz) the error in the L^∞ norm is smaller or approximately equal to h. In the last column of the table we show the quantity $(J(u) - J_h(u_h))/h$ which is approximately 0.14 for all the values of h reported, showing that even under such low regularity assumptions on u, the error in the functional behaves as $O(h)$.

In three dimensions there is also a solution with partial derivatives which are either 0 or 1, of the form

$$u(x, y, z) = \max \{0, x - a, y - a, z - a,$$
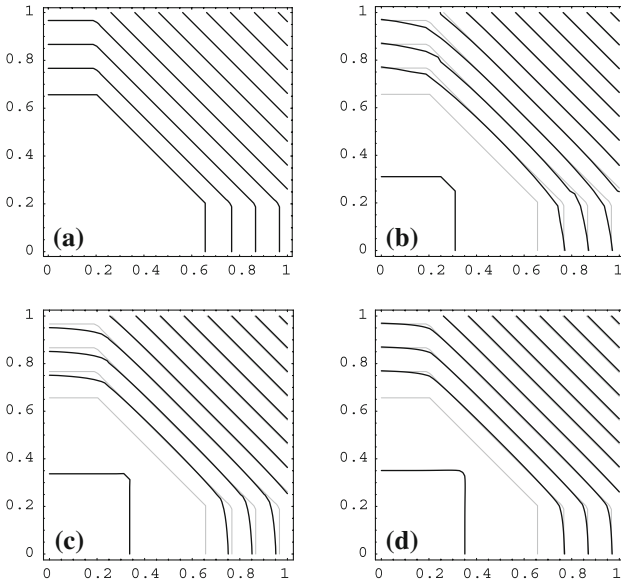$$x + y - b, x + z - b, y + z - b, x + y + z - c\}.$$

Fig. 1 *Contour lines* of the solutions of the economics problem in 2D: analytic solution (**a**), and discrete approximations with $h = 1/16$ (**b**), $h = 1/32$ (**c**), $h = 1/64$ (**d**)

Table 1 Comparison of the approximations to the economics problems in 2D for different mesh sizes

n	$h = 1/n$	$J_h(u_h)$	$J_h(I_h u)$	Error (L^∞)	Time (s)	$\frac{J(u) - J_h(u_h)}{h}$
8	0.125	0.5319	0.5444	0.0769	0.190	0.14
16	0.0625	0.5404	0.5478	0.0300	0.990	0.14
32	0.03125	0.5449	0.5488	0.0336	17.000	0.14
64	0.015625	0.5470	0.5491	0.0174	751.100	0.14
∞	0	0.5492				

The column labeled n denotes the number of subdivisions of the interval $[0, 1]$ in each direction. The number of unknowns is $(n + 1)^2$

This time the coefficients cannot be expressed in a simple form, since they involve roots of polynomials of high degree, and we just give numerical approximations:

$$a = 0.840627, \quad b = 1.038352, \quad c = 1.236077,$$

so that $b - a = c - b$, and the value of the functional is

$$J(u) \approx 0.868405.$$

In Fig. 2 we show the regions where $\nabla u \cdot (1, 1, 1)$ is 0, 1, 2 or 3 in, different shades of gray. We illustrate the regions with wire frames viewed from the positive octant in (a), and exploded views of the solid regions from the positive octant in (b) and the negative octant in (c).
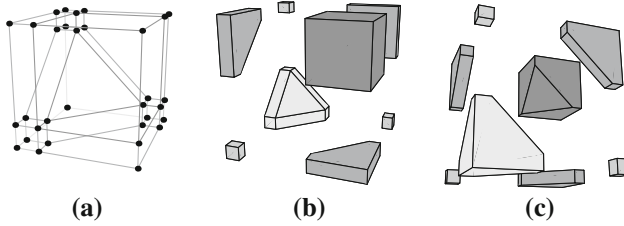
Fig. 2 Regions where the analytic solution in 3D has constant gradient: with wire frames viewed from the positive octant (**a**), and exploded views of the solid regions from the positive octant (**b**), and the negative octant (**c**)

Table 2 Comparison of the approximations to the economics problems in 3D for different mesh sizes

n	$h = 1/n$	$J_h(u_h)$	$J_h(I_h u)$	Error (L^∞)	Time (s)	$\frac{J(u) - J_h(u_h)}{h}$
4	0.25	0.8195	0.8449	0.1356	0.21	0.20
8	0.125	0.8484	0.8605	0.1281	3.87	0.16
12	0.0833	0.8578	0.8647	0.1130	10.29	0.13
16	0.0625	0.8622	0.8661	0.1135	7 m 20	0.10
20	0.0500	0.8648	0.8671	0.1177	50 m 23	0.07
∞	0	0.8684				

As before, the column labeled n denotes the number of subdivisions of the interval $[0, 1]$ in each direction. The number of unknowns is $(n + 1)^3$

In Table 2 we give some comparative results at different mesh sizes. It is interesting to notice here that the L^∞ error is not converging to zero with order $O(h)$. Nevertheless, the quantity $(J(u) - J_h(u_h))/h$ (shown in the last column) decreases slowly as h goes to 0, meaning that the error $J(u) - J_h(u_h)$ in the functional behaves as $O(h)$ in this range, exhibiting a similar behavior to that of the two dimensional case.

Remark 6.2 Theoretically, semidefinite programs are polynomial time solvable. However, as the Tables 1 and 2 show, in practice we cannot go too far with the number of unknowns.

A reasonable size for the two dimensional problems we considered is a mesh of about $n = 40$ subdivisions.
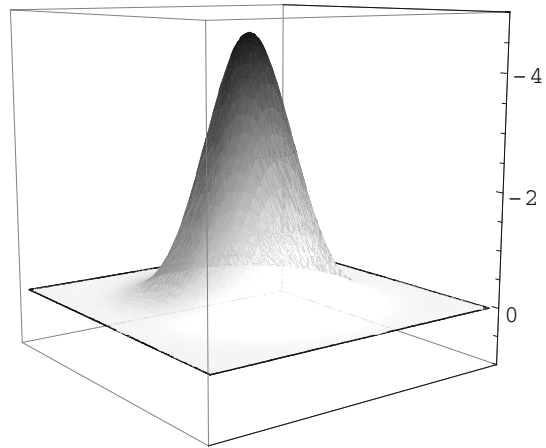
6.2 Projections

Carlier, Lachand-Robert and Maury [6] gave several examples of H^1 and H_0^1 projections, and in this section we consider one of the functions they considered, namely,

$$f(x_1, x_2) = -(4 + 5x_1 x_2^2) e^{-30\left((x_1 - 1/2)^2 + (x_2 - 1/2)^2\right)} \quad \text{for } (x_1, x_2) \in Q.$$

We show the graph of the original function in Fig. 3, and that of the resulting L^1, L^2, L^∞, H^1 and H_0^1 projections in Fig. 4. As in the original article, these graphs are

Fig. 3 Graph of
$f(x_1, x_2) = -(4 + 5x_1 x_2^2)$
$e^{-30[(x_1-1/2)^2+(x_2-1/2)^2]}$,
shown upside down

shown upside down. The interested reader may observe that our results are qualitatively different from those in [6, p. 304].

Using a similar function, but with more symmetries,

$$g(x_1, x_2) = -\sin^3(2\pi x_1)\, \sin^3(\pi x_2) \qquad \text{for } (x_1, x_2) \in Q,$$

we show in Fig. 5 that the L^∞ projection need not be unique. The solution in (b) is the one obtained by the semidefinite programming code.

In all these examples, we used a mesh with 41×41 nodes. The times ranged from about 40 seconds for the L^∞ projections to about 200 seconds for the L^2 projection. The H^1 and H_0^1 projections took about the same time, near 140 seconds.

6.3 Fitting data

As mentioned in the introduction, many problems in science are modelled via convex functions, raising the question of how measured data (usually non convex) can be approximated by a convex function.

Often the fitness to data is done parametrically. For example, by assuming that the underlying function is a linear combination of some given polynomials, and minimizing over all possible parameters in a convenient norm. Even in this case, approximating the data by a linear combination which is also convex—but otherwise arbitrary—may be challenging.

In this section we show how our numerical scheme could be used for fitting data on a regular mesh by discrete functions having a positive semidefinite discrete Hessian. Though the resulting discrete functions may not be extended to a convex function of continuous variables, as shown in Example 2.3, the underlying "true" convex function might be well captured by the discrete function.
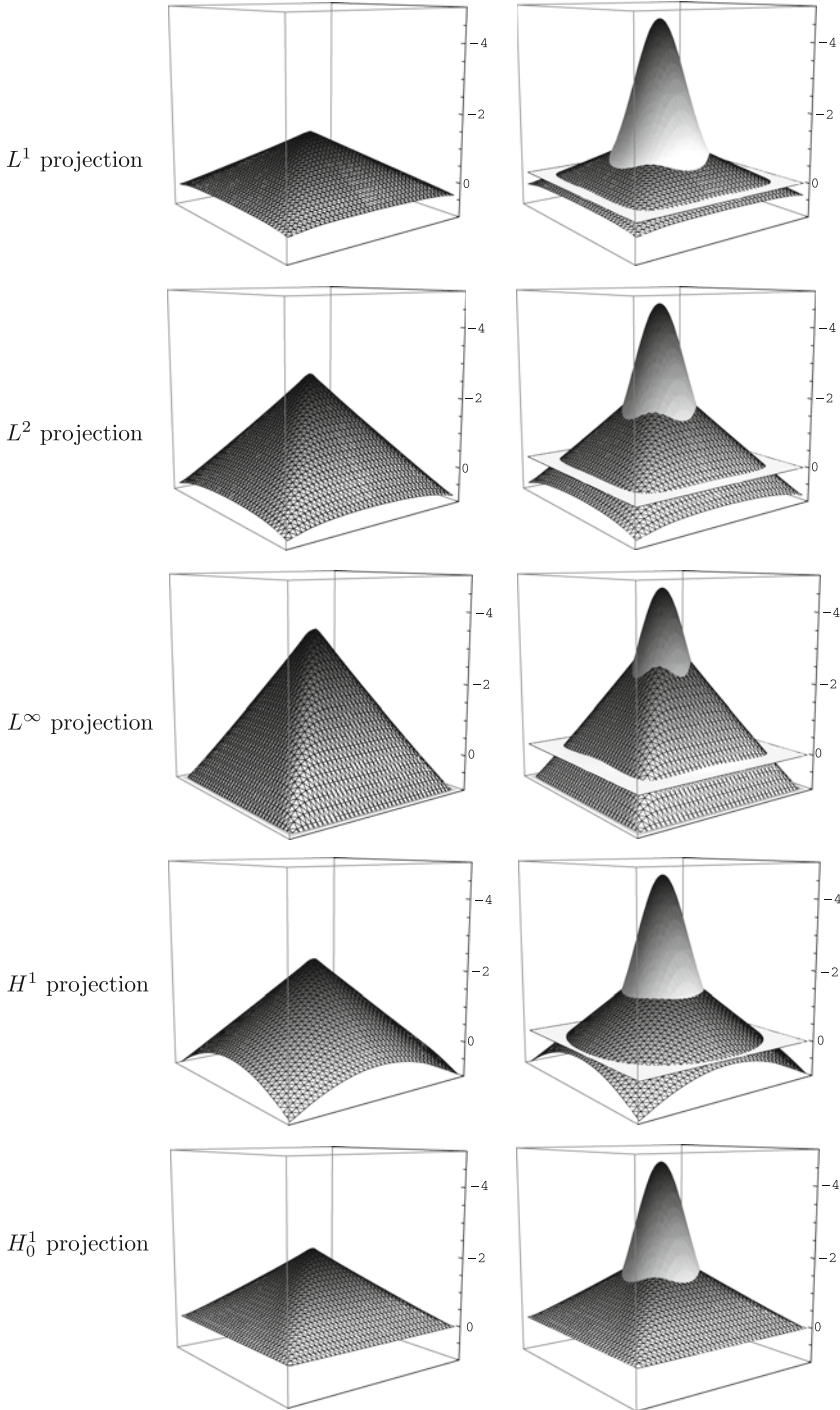
L^1 projection

L^2 projection

L^∞ projection

H^1 projection

H_0^1 projection

Fig. 4 Projections and its comparisons to the function of Fig. 3, shown upside down
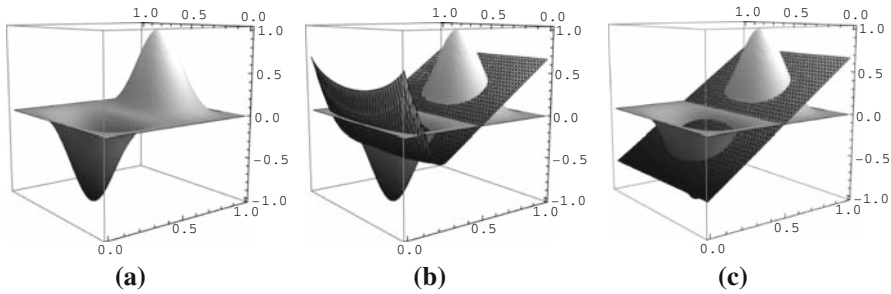
Fig. 5 Non-uniqueness of the L^∞ projection. **a** original function, **b, c** two optimal solutions

In the tests we show, we perturbed with random noise the values of the function

$$f(x_1, x_2) = (x_1 - 1/2)^2 + 2(x_2 - 1/2)^2,$$

on a regular mesh of 41×41 nodes. The original function and the perturbed data are represented in Fig. 6a and b, respectively.[1]

We chose to use a uniformly distributed noise between $-\varepsilon$ and ε for each node, where $\varepsilon = 10\,h$ (h being the mesh size). In this way we simulate some intrinsic measurement noise whose distribution is presumably known, and that the mesh has been chosen finer than the measurement noise. Needless to say, our choice of noise is quite arbitrary, and there are many other possibilities to choose from.

In Fig. 6c–e we show the resulting level curves for the L^1, L^2 and L^∞ discrete projections (respectively), with the contours of the original unperturbed function in a lighter gray. The L^2 discrete projection may be considered as a variant of a least squares approximation, but the others are not seen as often.

The times spent by the semidefinite code were similar to those of Sect. 6.2, about 170, 200 and 40 s (respectively). As is to be suspected, for the same mesh but smaller noise (e.g., taking $\varepsilon = h$), the results are better and the times smaller.

Comparing the level curves with those of the unperturbed function (Fig. 6c–e), it is apparent that the L^∞ discrete projection seems to give the closest fit. Notice that this projection is also faster than the L^1 or L^2 discrete projections by a factor of 4–5.

For the run shown, the maximum absolute value of the differences between the perturbed data and the original function on the mesh is about 0.25 (10 times the mesh size, as designed), whereas the maximum absolute value of the differences between the L^∞ projection and the original function is 0.37, attained at $(0, 0)$.

However, as can be seen in Fig. 6e, the L^∞ projection gives a good fit to the unperturbed function in the interior nodes, and is actually between ± 0.015 (somewhat smaller than the mesh size) for nodes of the square $0.1 \le x_1, x_2 \le 0.9$, but deteriorates near the boundary (as do the other projections). This is perhaps better appreciated in Fig. 6f, where the graph of the L^∞ discrete projection is compared to that of the unperturbed original function.

---

[1] It must be kept in mind that, though the data is discrete, the graphic software makes its own interpolations for drawing surfaces and level curves.
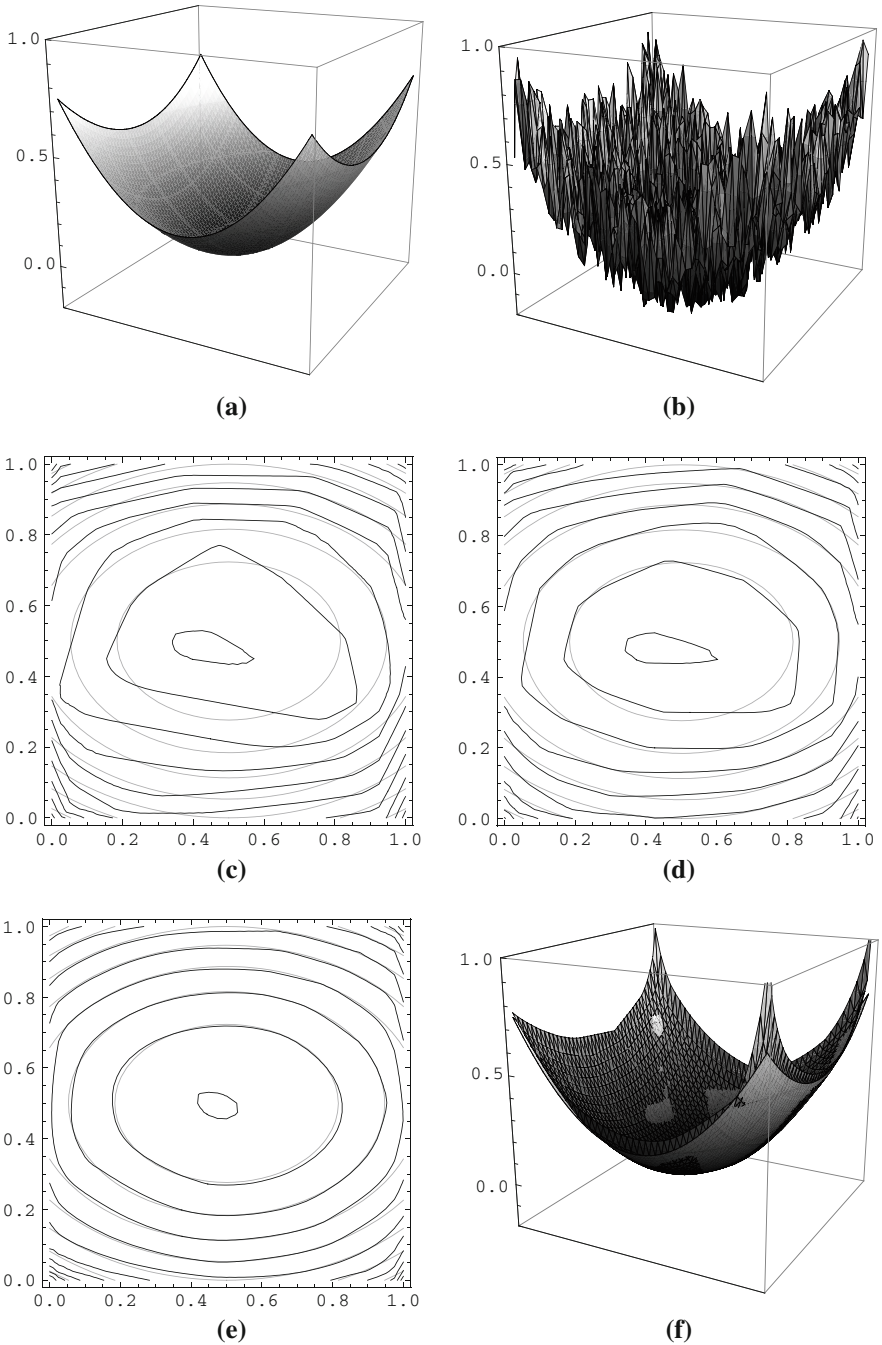
Fig. 6 Fitting a convex surface to perturbed data. Original function (**a**), perturbed data (**b**), level curves of the L^1 projection (**c**), level curves of the L^2 projection (**d**), level curves of the L^∞ projection (**e**), 3D graphs of the unperturbed function and L^∞ projection (**f**)

The "overshooting" at the boundary is a typical and known phenomenon when fitting data by convex functions. See, for instance, the article by Meyer [15], where the one dimensional case is discussed.

References

1. Birke, M., Dette, H.: Estimating a convex function in nonparametric regression. Scand. J. Stat. **34**(2), 384–404 (2007)
2. Borchers, B.: CSDP, A C library for semidefinite programming. Optim. Methods Softw. **11**(1), 613–623 (1999)
3. Buttazzo, G., Guasoni, P.: Shape optimization problems over classes of convex domains. J. Convex Anal. **4**(2), 343–351 (1997)
4. Buttazzo, G., Ferone, V., Kawohl, B.: Minimum problems over sets of concave functions and related questions. Math. Nachr. **173**, 71–89 (1995)
5. Carlier, G., Lachand-Robert, T.: Regularity of solutions for some variational problems subject to a convexity constraint. Commun. Pure Appl. Math. **LIV**, 583–594 (2001)
6. Carlier, G., Lachand-Robert, T., Maury, B.: A numerical approach to variational problems subject to convexity constraint. Numer. Math. **88**(2), 299–318 (2001)
7. Eaton, J.W.: GNU octave manual, Network Theory Limited (2002) http://www.octave.org
8. Groeneboom, P., Jonglbloed, G., Wellner, J.A.: Estimation of a convex function: characterizations and asymptotic theory. Ann. Stat. **29**(6), 1653–1698 (2001)
9. Hanson, D.L., Pledger, G.: Consistency in concave regression. Ann. Stat. **4**(6), 1038–1050 (1676)
10. Hoffman, A.J., Wielandt, H.W.: The variation of the spectrum of a normal matrix. Duke Math. J. **20**, 37–39 (1953)
11. Lachand-Robert, T., Oudet, É.: Minimizing within convex bodies using a convex hull method. SIAM J. Optim. **16**(2), 368–379 (2005)
12. Lachand-Robert, T., Peletier, M.A.: Extremal points of a functional on the set of convex functions. Proc. Am. Math. Soc. **127**(6), 1723–1727 (1999)
13. Lachand-Robert, T., Peletier, M.A.: Newton's problem of the body of minimal resistance in the class of convex developable functions. Math. Nachr. **226**, 153–176 (2001)
14. Manelli, A.M., Vincent, D.R.: Multidimensional mechanism design: revenue maximization and the multiple-good monopoly. J. Econ. Theory **137**, 153–185 (2007)
15. Meyer, M.C.: Consistency and power in tests with shape-restricted alternatives. J. Stat. Plan. Inference **136**, 3931–3947 (2006)
16. Murota, K., Shioura, A.: Relationship of M-/L-convex functions with discrete convex functions by Miller and by Favati-Tardella. Discrete Appl. Math. **115**, 151–176 (2001)
17. Rochet, J.-C., Choné, P.: Ironing, sweeping and multidimensional screening. Econometrica **66**, 783–826 (1998)
18. Shih, T.D., Chen, V.C.P., Kim, S.B.: Convex version of multivariate adaptive regression splines for optimization (2006). http://students.uta.edu/dt/dts5878/convexMARS.pdf
19. Rockafellar, R.T.: Convex analysis. Princeton University Press, Princeton (1970)
20. Vandenberghe, L., Boyd, S.: Semidefinite programming. SIAM Rev. **38**(1), 49–95 (1996)
21. Wolfram Research, Inc.: Mathematica, version 6.0. Wolfram Research, Inc (2007)
22. Ziemer, W.P.: Weakly differentiable functions: Sobolev spaces and functions of bounded variation. Springer, Heidelberg (1989)