Journal of Mathematical Imaging and Vision 12, 99–108 (2000)

Ambiguous Shape from Shading with Critical Points

JEAN-DENIS DUROU
Institut de Recherche en Informatique, Université Paul Sabatier (Toulouse III), 118, route de Narbonne, 31062 Toulouse Cedex, France
durou@irit.fr

DIDIER PIAU
Laboratoire de Probabilités, Université Claude Bernard (Lyon I), 43, boulevard du 11 Novembre 1918, 69622 Villeurbanne Cedex, France
piau@jonas.univ-lyon1.fr

Abstract. The eikonal equation can have an infinite number of solutions when the image has critical points. We exhibit a family, indexed by a continuous parameter, of non isomorphic surfaces with one critical point, which give the same simple image. Hence, shape from shading can be an ill-posed problem when no additional condition on the shape is imposed, even when the image has critical points. Also, deformations without distortion are possible, i.e., there can exist a continuous deformation of the surface which does not modify its image.

Keywords: shape from shading, eikonal equation, ill-posed problem, shape reconstruction

1. Introduction

Given a single image of a surface S, one can determine the shape of S through its texture, shadows, contours or shading. If S is non textured, has no edge nor hidden part, and if no silhouette is visible in the image, one has to rely on the shading of the image.

The study of these so-called *shape from shading* methods, pioneered by Horn in the 1970s, see [5], relies on the following *image irradiance equation*, see [6]:

$$R(\nabla f(x, y)) = E(x, y), \tag{1}$$

where the image is described through the shading $E(x, y)$ of each point (x, y), and the shape of S is defined by the equation $z = f(x, y)$. The zOz' axis points towards the observer and $f(x, y)$ is the altitude of S above point (x, y). Finally, the re-emission of light by S at point (x, y) in the direction of observation is a known function $R(\nabla f)$ of the gradient

$$\nabla f = (\partial_x f, \partial_y f).$$

Assuming that the light source is unique, infinitely distant from the surface, and that it comes from the direction of the observer, i.e., from the zOz' axis, Lions, Rouy and Tourin [8] show that the shading $R(\nabla f)$ of the image depends only on the slope of S, as measured by the Euclidean norm of the gradient. Hence,

$$R(\nabla f) = r((\partial_x f)^2 + (\partial_y f)^2). \tag{2}$$

For example, when the surface is Lambertian, r is given by $r(u) = r_0/(1 + u)^{1/2}$, where r_0 denotes a positive constant. Notice that, if the light source comes from the direction of the point $(x_0, y_0, 1)$ and if the surface is still Lambertian, (2) should be replaced, see [8], by

$$R(\nabla f) = r_0(1 - x_0\, \partial_x f - y_0\, \partial_y f)/(1 + u)^{1/2},$$

where $u = (\partial_x f)^2 + (\partial_y f)^2$.

Confining ourselves for the rest of this paper to a unique light source coming from the zOz' axis, one can show that, for usual materials, the function r is decreasing. This implies that r^{-1} exists, and that (1) can be rewritten as:

$$(\partial_x f)^2 + (\partial_y f)^2 = e(x, y), \tag{3}$$

where $e(x, y) = r^{-1} \circ E(x, y)$ is supposedly known. The *eikonal equation* (3) is a non linear first order partial differential equation and it has been widely studied, see [2, 3, 9]. Of major interest is to know whether (3) has a finite number of non isomorphic solutions (i.e., solutions which describe really different shapes) or not, in the absence of boundary conditions, that is if one does not a priori know the altitude of any point of the image.

In [2, 7], eikonal equations which admit no continuously differentiable solution are exhibited. On the other hand, conditions on the right hand side function $e(x, y)$ are given in [1, 3, 9], which ensure that (3) admits a unique continuously differentiable solution. These uniqueness results assume the existence of a *silhouette* in the image, i.e., of a closed curve along which $e(x, y)$ is infinite. Other uniqueness results with similar conditions are given in [8], using viscosity solutions. Such situations are interesting because they fulfill the initial wish of Horn in [5] that the image be unambiguous. However, shape from shading, seen as the inverse problem of the formation of images, can be ill-posed, i.e., there can exist an infinite family of surfaces which give the same right hand member $e(x, y)$, even in the presence of critical points (in our context, a critical point is a point (x, y) where $e(x, y) = 0$).

The accepted wisdom with respect to the ill-posedness of shape from shading seems to be that the construction of a family (f_t) of solutions of (3) is straightforward when no critical point exists. Starting from a solution f_0, the reasoning in this case proceeds as follows. Consider the family of curves of equations $f_t(\cdot) = h$, for all $h \in \mathbf{R}$, and the family of curves of equations $\partial_t f_t(\cdot) = h$, for all $h \in \mathbf{R}$. Assume that these two families are orthogonal. Then,

$$\nabla f_t(x, y) \cdot \nabla \partial_t f_t(x, y) = 0, \tag{4}$$

and this ensures that each f_t is a solution of (3). (To see this, differentiate (3) with respect to t, getting (4).) Of course, in order to make a proof from this heuristics, one would still have to show that the differential Eq. (4) could indeed be integrated. Furthermore, all the solutions which are built with this procedure could represent isomorphic surfaces. (Consider $e(x, y) = 1$ and $f_0(x, y) = x$. From (4), one can get for example $f_t(x, y) = x \cos(t) - y \sin(t)$, and f_t is the composition of f_0 and of a rotation.)

In order to define precisely when two solutions are isomorphic, notice that, if S defined by $z = f(x, y)$ is a solution of (3), then $S + c$ and $-S + c$, defined respectively by $z = f(x, y) + c$ and $z = -f(x, y) + c$, are also solutions, for any real number c. Moreover, for any rotation or symmetry Θ of the (x, y) plane, the surface defined by $z = f \circ \Theta^{-1}(x, y)$, has the same shape as S. These remarks motivate the following definition.

Definition 1. The surfaces S and S' are isomorphic if there exists an orthogonal linear transformation Θ of the (x, y, z) space and a point s_0 such that $S' = \Theta(S) + s_0$, i.e., such that $(x, y, z) \in S$ if and only if $\Theta(x, y, z) + s_0 \in S'$.

The surfaces S and S' are isomorphic solutions of (3) if S and S' are solutions of (3) and if they are isomorphic.

More important than this problem of isomorphic solutions, the heuristics above can fail when the image contains critical points. Assume for instance that the contour lines of f_0 are circles of center $(0, 0)$. The contour lines of $\partial_t f_0$ should be straight lines passing by $(0, 0)$, so what could be the value of $\partial_t f_0$ at this point? Finally, let us stress once again that, with or without critical points, the actual integration of (4) may be far from trivial to perform.

2. Results

We prove in this paper that eikonal equations (3) with critical points can admit a continuous one-parameter family of non isomorphic solutions, by exhibiting a simple example. (Recall that, in our context, a critical point is a point (x, y) where $e(x, y) = 0$ in (3) and that isomorphic surfaces are defined in Section 1.) To this end, consider the elliptic surface S_E defined by the equation $z = 2x^2 + y^2$, see Fig. 1.

Figure 2 shows the image of S_E, computed with the MATLAB 4.2 software, under the hypotheses of front lighting and of diffuse reflection.

We prove that S_E is rigid with respect to shape from shading, i.e., that S_E cannot be continuously deformed without distortion (see Proposition 1 of Section 3 for a precise statement). On the other hand, S_E gives the same image as the hyperbolic surface S_H defined by the equation

$$(S_H) \quad z = 2x^2 - y^2,$$

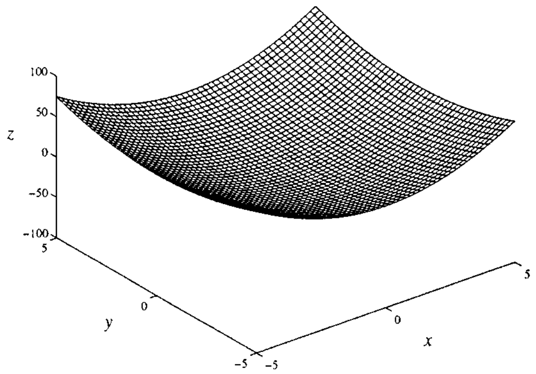

Figure 1. Surface S_E: $z = 2x^2 + y^2$.

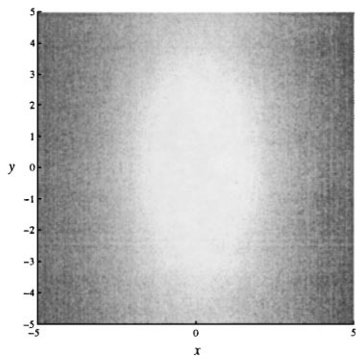

Figure 2. Image of S_E lit from the direction of the observer.

since their common associated function e is

$$e(x, y) = 16x^2 + 4y^2.$$

Theorem 1 shows that S_H is not rigid with respect to shape from shading.

Theorem 1. *Consider the eikonal equation*

$$(\partial_x f)^2 + (\partial_y f)^2 = 16x^2 + 4y^2. \qquad (5)$$

There exists a family (S_t), *indexed by* $t \in]0, +\infty[$, *of surfaces* S_t *defined by the functions* $f_t : \mathbf{R}^2 \to \mathbf{R}$ *such that the following properties hold*:

(i) *Each* f_t *is real analytic on the open disc* B_t *of center* $(0, 0)$ *and radius* t.
(ii) *Each* f_t *is a solution of* (5) *on* B_t.
(iii) *The function* $(t, x, y) \mapsto f_t(x, y)$ *is continuous on* $]0, +\infty[\times \mathbf{R}^2$.
(iv) *For* $t \neq u$, S_t *and* S_u *are not isomorphic.*
(v) *When t goes to infinity,* f_t *converges uniformly on every compact domain to* f_H, *which corresponds to the surface* S_H.

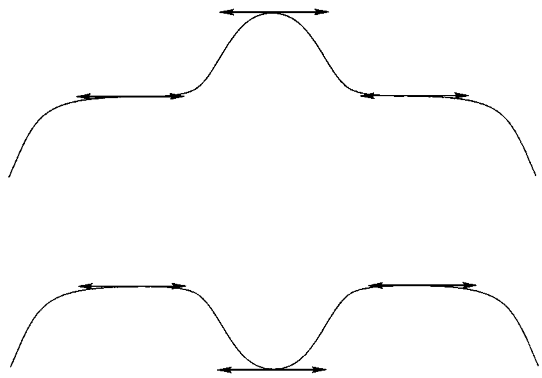

Figure 3. Two non isomorphic surfaces which yield the same image.

Each surface S_t in Theorem 1 has a critical point at $(0, 0)$, which is a saddle point. According to part (i), each S_t is as smooth as possible, i.e., real analytic, in a neighbourhood of $(0, 0)$. According to (ii) and (iii), the deformation $t \mapsto S_t$ itself is invisible and continuous (and in fact real analytic where the surfaces are smooth). Part (v) shows that the family (S_t), at least for $t \geq u$, is indeed a smooth and invisible deformation of S_H when S_H is observed in a bounded window B_u. Note that part (iv) is also true when the surfaces are restricted to any bounded window around any point of the (x, y) plane. Finally, we point out that images with multiple critical points can in a natural way yield a large number of non isomorphic solutions. Figure 3 plots two surfaces of a simple example due to [8], which obviously yield the same image. In Theorem 1, the critical point is unique.

Remark. We mention that the result of this paper can be carried out for the following circularly symmetric eikonal equation:

$$(\partial_x f)^2 + (\partial_y f)^2 = 4x^2 + 4y^2.$$

In Section 3, we look for solutions of (5), writing them as power series expansions around the origin $(0, 0)$. This yields the rigidity of S_E, namely that the only solution which is locally convex (resp. concave) at $(0, 0)$ is S_E (resp. $-S_E$). On the contrary, solutions which have a saddle point at $(0, 0)$ are not uniquely determined. In Section 4, we exhibit a family (f_t) which satisfy the properties stated in Theorem 1. Each f_t has a saddle point at $(0, 0)$. Further properties of (f_t) are mentioned at the beginning of Section 4.

3. Power Series Solutions of (5)

Let S be the surface defined by $z = f(x, y)$ where f is a solution of (5). Set for instance $f(0, 0) = 0$ and assume that f is real analytic at $(0, 0)$, i.e., that f is the sum of a power series in x and y in a neighbourhood of $(0, 0)$. Hence,

$$f(x, y) = \sum_{n+p\geq 1} a_{n,p} x^n y^p \tag{6}$$

for $x^2 + y^2$ small enough. One can compute from (6) the first terms of the power series expansion of $\|\nabla f\|^2$. Identifying the terms of degree ≤ 2 with the right hand side of (5), one gets $a_{1,0} = a_{0,1} = a_{1,1} = 0$ and

$$a_{2,0} = \pm 2, \quad a_{0,2} = \pm 1.$$

Turning to the terms of degree 3, one gets $a_{3,0} = a_{2,1} = a_{0,3} = 0$, while

$$a_{1,2}(a_{2,0} + 2a_{0,2}) = 0.$$

Taking into account the possible values of $a_{2,0}$ and $a_{0,2}$, two cases arise.

- *Hyperbolic case*: $a_{2,0}$ and $a_{0,2}$ have opposite signs. Then, S has a saddle point at $(0, 0)$ and $a_{1,2}$ is undetermined. This case yields the family (f_t) of Section 4.
- *Elliptic case*: $a_{2,0}$ and $a_{0,2}$ have the same sign. Then, S is locally convex or locally concave at $(0, 0)$ and $a_{1,2} = 0$. Proposition 1 settles the study of this elliptic case.

Proposition 1. *The only power series solutions of* (5) *such that* $a_{2,0}$ *and* $a_{0,2}$ *have the same sign are* $f_E(x, y) = 2x^2 + y^2$ *and* $-f_E$. *Hence, the only surfaces which are solutions of* (5), *up to isomorphism, are* S_E *and surfaces with a saddle point at* $(0, 0)$.

Proof: All the terms of degree 3 are zero. Assume that all the terms of degree between 3 and $n - 1$ are zero and, without loss of generality, that $a_{2,0} = 2$ and $a_{0,2} = 1$. Then,

$$f(x, y) = 2x^2 + y^2 + \sum_{p=0}^{n} a_{p,n-p} x^p y^{n-p} + o((x^2 + y^2)^{n/2}). \tag{7}$$

Differentiating (7) with respect to x and to y and keeping only the terms of degree n in (5), one gets

$$\sum_{p=0}^{n} a_{p,n-p}(2xpx^{p-1}y^{n-p} + y(n - p)x^p y^{n-p-1}) = 0.$$

Thus, for any p, $(2p + (n - p))a_{p,n-p} = 0$, which yields $a_{p,n-p} = 0$ since $(2p + (n - p)) = n + p \neq 0$. This proves the recursion. □

Remark. The proof above shows why the hyperbolic case is undetermined. If $a_{2,0} = 2$ and $a_{0,2} = -1$, one would get the condition

$$(2p - (n - p))a_{p,n-p} = 0.$$

Hence, the coefficient of each $x^p y^{2p}$ can be chosen at will. We mention that the parameter t of the family (f_t) of solutions given in Section 4 could be defined as

$$a_{1,2} = 2/t.$$

Also, the family of Section 4 by no means exhausts the set of solutions which should be indeed parametrized by all the coefficients $a_{p,2p}$ for $p \geq 1$.

4. The Hyperbolic Case: A Family of Solutions

We now exhibit a family (f_t) of solutions of (5) of Theorem 1. We will prove the following additional properties:

(vi) *The function* $(t, x, y) \mapsto f_t(x, y)$ *is real analytic on* $(]0, +\infty[\times \mathbf{R}^2)\backslash D$, *with*

$$D = \{(t, x, y);\ y = 0, x + t \leq 0\}.$$

(vii) *For a fixed* $t \in]0, +\infty[$, f_t *is real analytic on* $\mathbf{R}^2\backslash D_t$, *with*

$$D_t =]-\infty, -t] \times \{0\}.$$

(viii) *When t converges to zero, f_t converges uniformly on every compact domain to f_0, defined by*

$$f_0(x, y) = 2x|x| + y^2.$$

Finally, we mention without giving the proof here the fact that, at $P_t = (-t, 0)$, f_t is continuously differentiable, but not twice continuously differentiable, and that, at any point of $D_t \backslash \{P_t\}$, f_t is continuous but not differentiable.

4.1. Construction of f_t

We define f_t by the relation:

$$f_t(x, y) = 2x^2 + y^2 + \frac{4}{3} g(a, b)(2a - g(a, b)), \quad (8)$$

where a, b and $g(a, b)$ are given below. Introduce the polynomial function H defined by

$$H(g, a, b) = (g - 2a)(g + a)^2 - 2b. \quad (9)$$

Since $H(\cdot, a, b)$ is a polynomial of degree 3, it has always at least one real root. For $(a, b) \in \mathbf{R} \times \mathbf{R}^+$, we define a real number $g(a, b)$ by any of the two following properties:

Continuity definition: $g(a, b)$ is a root of $H(\cdot, a, b)$, the function $(a, b) \mapsto g(a, b)$ is continuous and $g(0, \frac{1}{2}) = 1$.
Greatest root definition: $g(a, b)$ is the greatest real root of $H(\cdot, a, b)$.

We prove in Section 4.2 that these two definitions are equivalent. We complete the definition of $f_t(x, y)$ by setting

$$a = x + t, \quad b = \frac{27}{8} t y^2.$$

Remark. Although the aspect of (8) could lead to think otherwise, each f_t has indeed a saddle point at $(0, 0)$. One can prove that

$$f_t(x, y) = 2x^2 - y^2 + 2xy^2/t + o((x^2 + y^2)^{3/2}).$$

Remark. The definition of the functions f_t may seem arbitrary. The way by which we get to these expressions involves heavy computations and does not seem to us to be interesting *per se*. Hence, we skip this part in the present exposition.

Equation (8) gives $f_t(x, y)$ as a polynomial function of $(x, y, t, g(a, b))$. Furthermore, (a, b) is polynomial in (x, y, t). Hence, the continuity, respectively the real analyticity, of $g(a, b)$ with respect to (a, b) implies the same regularity of $f_t(x, y)$ with respect to (x, y). We now study the regularity properties of $g(a, b)$.

4.2. Explicit Definition of $g(a, b)$

Set $r = a^3 + b$. The discriminant, see [4], of the polynomial $H(\cdot, a, b)$ is

$$s = r^2 - a^6 = b(2a^3 + b),$$

and the sign of s gives the number of real roots of $H(\cdot, a, b)$. Figure 4 shows the corresponding decomposition of the (a, b) upper half plane $\mathbf{R} \times \mathbf{R}^+$.

- On $A^+ = \{s > 0\}$, $H(\cdot, a, b)$ has one real root, given by Cardan's formula:

$$g(a, b) = (r + s^{1/2})^{1/3} + (r - s^{1/2})^{1/3}. \quad (10)$$

  Since $r > 0$ on this domain, $s^{1/2} - r < s^{1/2} + r$ and $g(a, b) > 0$. Finally, for $(a, b) = (0, \frac{1}{2})$, one has $r = \frac{1}{2}$, $s = \frac{1}{4}$ and $g = 1$.
- On $A^- = \{s < 0\}$, $H(\cdot, a, b)$ has three real roots, given by

$$g_k = -2a\cos(\phi_k),$$

  where $\phi_2 - \phi_1 = \phi_3 - \phi_2 = 2\pi/3$ and

$$\phi_1 = \frac{1}{3}\arccos(-r/a^3). \quad (11)$$

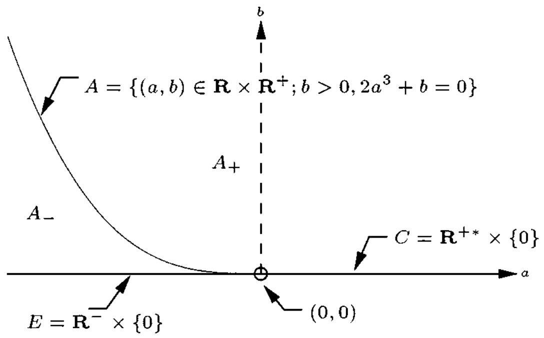

Figure 4. Decomposition of the (a, b) half plane.

Notice that the formula (11) makes sense on A^- since $s < 0$ implies $b > 0$ and $2a^3 + b < 0$, hence

$$a^3 < a^3 + b = r < -a^3.$$

Recall that, by definition, $\arccos(\phi)$ belongs to $[0, \pi]$. Hence, $\phi_1 \in]0, \pi/3[$, $g_1 > 0$ and $g_2 < g_3 < g_1$. We set $g(a, b) = g_1$.

- On $A = \{s = 0, b > 0\}$, one has

$$H(g, a, b) = (g + 2a)(g - a)^2.$$

We set $g(a, b) = -2a > 0$.

- On $C = \{b = 0,\ a > 0\}$ and on $E = \{b = 0,\ a \leq 0\}$, one has

$$H(g, a, b) = (g - 2a)(g + a)^2.$$

We set $g(a, b) = 2a > 0$ on C and $g(a, b) = -a \geq 0$ on E.

The choices written above imply that, on the whole (a, b) upper half plane, one has:

$*$ *$g(a, b)$ is the greatest real root.*
$*$ *$g(a, b)$ is the only nonnegative root.*
$*$ *$g(a, b)$ is a simple root of $H(\cdot, a, b)$ if $(a, b) \notin E$.*
$*$ *$g(a, b)$ is positive, except when $(a, b) = (0, 0)$.*

Hence, our choices fulfill the "greatest root" definition. As for the "continuity" definition, g is continuous on A^+ and on A^-. In order to prove the continuity of g everywhere, let (a_0, b_0) be a point of A, C or E, and assume that (a, b) converges to (a_0, b_0). Let g_0 be any limit point of $g(a, b)$. Then, $g_0 \geq 0$ since $g(a, b) \geq 0$ everywhere. Furthermore, g_0 must be a root of $H(\cdot, a_0, b_0)$ by continuity of H. Hence, $g_0 = g(a_0, b_0)$ since $g(a_0, b_0)$ is the only nonnegative root of $H(\cdot, a_0, b_0)$. This proves that g is continuous at (a_0, b_0).

Figure 5 plots the real roots of $H(\cdot, a, b)$ at any point $(x, y) \in \mathbf{R}^2$, when $t = 1$.

In order to give a better understanding of the shape of this surface, Fig. 6 shows the section $x = -2$.

Figures 7 and 8 plot the functions g_1 defined by $g_1(x, y) = g(a, b)$ when $t = 1$, and f_1, respectively.

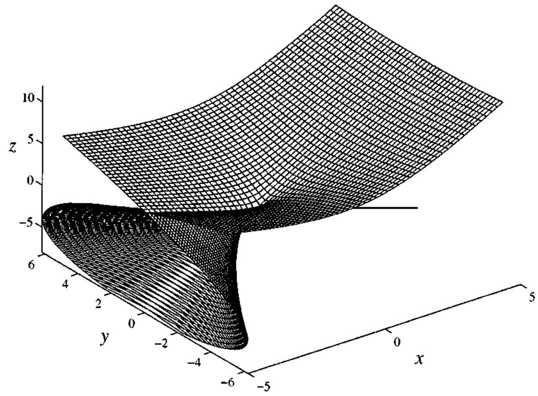

Figure 5. Real roots of $H(\cdot, a, b)$ for $t = 1$.

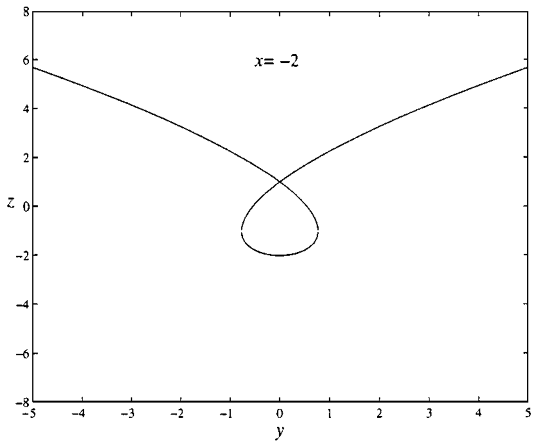

Figure 6. Section $x = -2$ of the surface of Fig. 5.

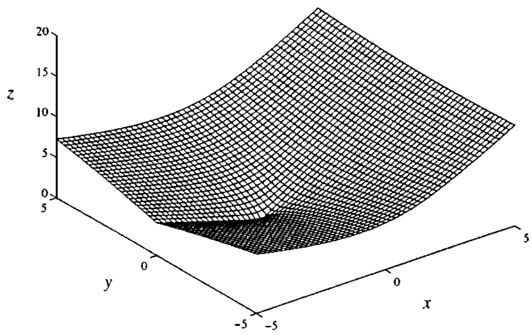

Figure 7. The function $(x, y) \mapsto g_1(x, y)$.

4.3. *Real Analyticity of g*

One sees that $g(a, b)$ is always a simple root of $H(\cdot, a, b)$, when $(a, b) \notin E$, and that g is continuous. This shows that g is in fact the local inversion of $H(\cdot, a, b)$ defined in a neighbourhood of the point $(0, \frac{1}{2})$. Recall that, at any point where g_0 is a simple root of $H(\cdot, a_0, b_0)$, the usual theorem of local inversion yields a function g, defined and continuous on a neighbourhood U of (a_0, b_0), such that $g(a_0, b_0) = g_0$

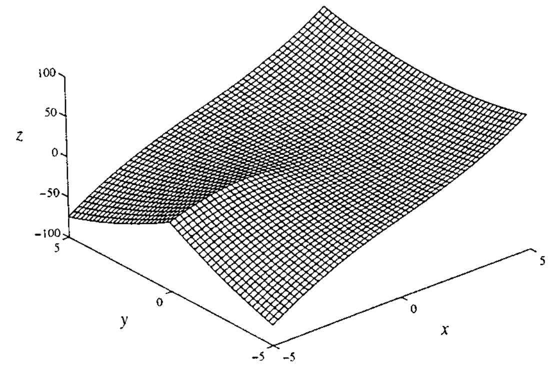

Figure 8. The function $(x, y) \mapsto f_1(x, y)$.

and $H(g(a, b), a, b) = 0$ on U. Furthermore, the local inversion theorem for analytic functions, see [4], ensures that g is in fact real analytic on U, where U can be any domain on which $g(a, b)$ is everywhere a simple root. This proves the real analyticity of g on $\{b > 0\}$, hence the real analyticity of f_t on $\{y \neq 0\}$, since the function $(a, b) \mapsto (x, y)$ is obviously real analytical at any (a, b) such that $b > 0$.

On the contrary, the half line $C = \{b = 0,\ a > 0\}$ needs a little extra care. Fix $(a_0, 0) \in C$. In a neighbourhood of $(a_0, 0)$, g is given by Cardan's formula (10), since (10) is valid on A^+ and, by continuity, on C as well. Furthermore,

$$s^{1/2} = \pm cyu^{1/2}, \quad c = (27t/8)^{1/2},$$

where the $\pm$ sign is given by the sign of y and where $u = 2a^3 + b$. In a neighbourhood of $(a_0, 0)$, r and u are positive and (10) may be re-written as

$$g(a, b) = (r + cyu^{1/2})^{1/3} + (r - cyu^{1/2})^{1/3}. \quad (12)$$

The right hand side of (12) is real analytic with respect to (a, y), hence $g(a, b)$ is real analytic with respect to (a, y), everywhere except on E. This implies that f_t is real analytic everywhere except on D_t.

Remark. The function g is also real analytic with respect to (a, b) at points of C different from $(0, 0)$ but we do not need to prove this stronger property.

Remark. A different proof of the analyticity of g on $\{b > 0\}$ entails defining g everywhere by Cardan's formula (10), and using well chosen determinations of the square and cubic roots functions on the complex plane.

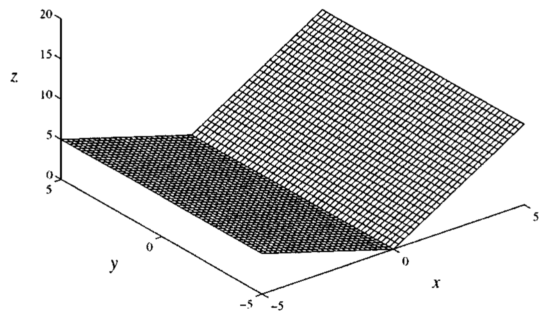

Figure 9. The function $(x, y) \mapsto g_0(x, y)$.

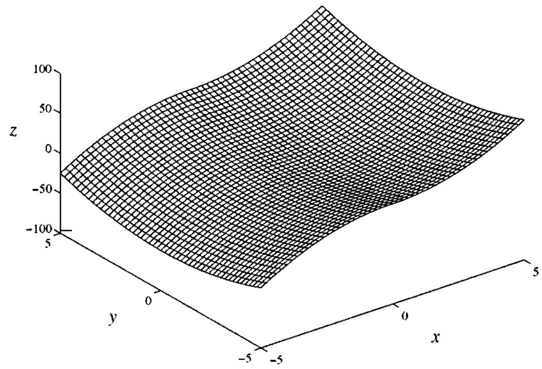

Figure 10. The function $(x, y) \mapsto f_0(x, y)$.

4.4. Limit Cases

4.4.1. The Case t = 0. When $t = 0$, the definitions above still make sense. The following choice of $g_0(x, y)$ guarantees that $g_t(x, y) = g(a, b)$ converges to $g_0(x, y)$ when t converges to zero. Here again, $g_0(x, y)$ must be nonnegative, hence

$$g_0(x, y) = 2x(x > 0), \quad g_0(x, y) = -x(x \leq 0).$$

Equation (8) then gives

$$f_0(x, y) = 2x|x| + y^2, \quad (13)$$

which proves property (viii), see Figs. 9 and 10.

4.4.2. The Set E. In this section, $t > 0$ and we study the regularity of f_t and g_t, defined by $g_t(x, y) = g(a, b)$ at points $(a, b) \in E$, or equivalently at points $(x, y) \in D_t$. We first prove that g_t is not differentiable. Recall that, for $a > 0$, i.e., for $x > -t$, one has

$$g_t(x, 0) = 2(x + t), \quad f_t(x, 0) = 2x^2,$$

while for $a \leq 0$, i.e., for $x \leq -t$, one has

$$g_t(x,0) = -(x+t), \qquad f_t(x,0) = 2x^2 - 4(x+t)^2.$$

We fix a point $(x,0) \in D_t$ and we look for a limited expansion of $g_t(x,y)$ when y is near zero. Using (9), one gets

$$g_t(x,y) = -a + c(x)y + o(y),$$

where $c(x)$ depends on y, only through its sign:

$$c(x) = \pm\frac{3}{2}(-t/a)^{1/2}.$$

We need the following Lemma.

Lemma 1. *If* $(a,b) \notin E$, *then* $g(a,b) > |a|$.

Proof: For any $(a,b) \notin E$, one has

$$H(|a|,a,b) = -2|a|^3 - 2a^3 - 2b < 0.$$

Since $H(g,a,b)$ goes to $+\infty$ when g goes to $+\infty$, this implies that $H(\cdot,a,b)$ has a real root which is greater than $|a|$. This proves Lemma 1 because $g(a,b)$ is the greatest real root of $H(\cdot,a,b)$. □

Lemma 1 implies that

$$g_t(x,y) = -a + \frac{3}{2}(-t/a)^{1/2}|y| + o(y).$$

Hence, g_t has a wedge at $(x,0) \in D_t$ and is not differentiable on D_t (even at $P_t = (-t,0)$ since, there, the slope of the wedge is infinite). Figures 11(a), (b) and (c), show three sections of g_1 through points of D_1: $x=-1$, $x=-3$ and $x=-5$.

Going back to f_t, (8) yields

$$f_t(x,y) = 2x^2 - 4a^2 - 8(-ta)^{1/2}|y| + o(y),$$

near a point $(x,0)$ of D_t. Hence, f_t is not differentiable on $D_t \backslash \{P_t\}$. The wedge of f_t vanishes at P_t, hence f_t is differentiable at P_t, where $\nabla f_t = (-4t,0)$. Finally, f_t cannot be real analytic at P_t since this would imply that f_t is differentiable in a neighbourhood of P_t. This proves properties (vi) and (vii). Figures 12(a), (b) and (c), show the three sections of f_1 which correspond to Fig. 11.

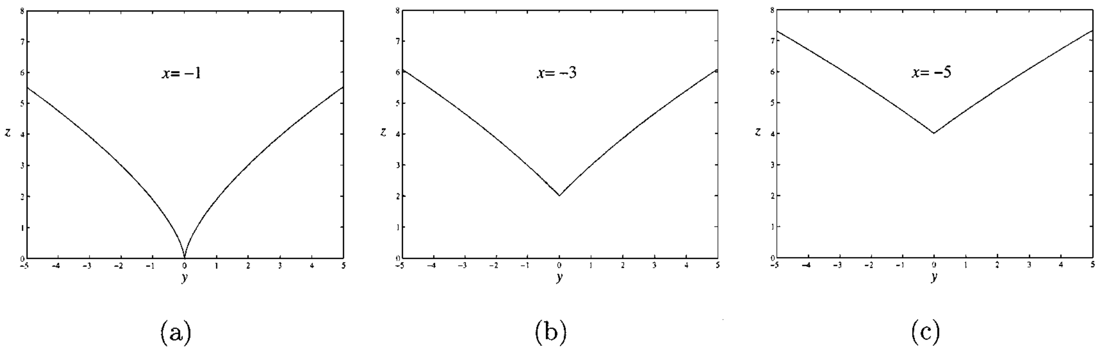

Figure 11. Three sections of g_1.

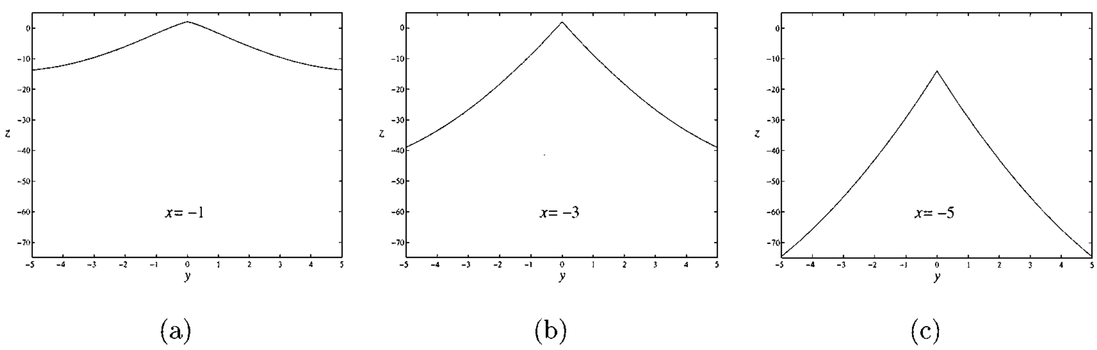

Figure 12. Three sections of f_1.

4.4.3. The Case $t = +\infty$. We prove property (v). Fix (x, y) and let t go to infinity. Then, $a \sim t$, so that (x, y) belongs to A^+ or C when t is large enough, and $g(a, b)$ is given by Cardan's formula (10). We look for an expansion of $g(a, b)$ with respect to t and to powers of $1/t$, using the equivalents:

$$r \sim a^3, \quad s \sim 2ba^3.$$

Expanding the cubic and square roots of (10), one gets

$$g(a, b) = 2a + \frac{2}{9} b/a^2 + o(1/t),$$

which, in terms of (x, y), is equivalent to

$$g_t(x, y) = 2(x + t) + \frac{3}{4} y^2/t + o(1/t).$$

Finally, using (8), the expansion of f_t reads as follows:

$$f_t(x, y) = 2x^2 - y^2 + o(1).$$

This means that S_t converges to the hyperbolic surface S_H, see Fig. 13. Notice that the point P_t goes to $(-\infty, 0)$ and that f_H is real analytic on $\mathbf{R}^2$.

4.5. f_t *is a Solution of* (5)

The expression (13) of f_0 shows that f_0 is solution of (5). Turning to the case $t > 0$, we differentiate $H(\cdot, a, b)$ with respect to x and to y. Using the expression (9) of $H(g, a, b)$, one gets

$$(g^2 - a^2)\, \partial_x g = 2ag + 2a^2,$$
$$(g^2 - a^2)\, \partial_y g = \frac{9}{2} ty.$$

When $(a, b) \notin E$, Lemma 1 ensures that $g \neq -a$. Hence, for $(a, b) \notin E$, one has

$$(g - a)\, \partial_x g = 2a, \tag{14}$$
$$(g - a)\, \partial_y g = \frac{9}{2} ty/(g + a). \tag{15}$$

The expression (8) of $f_t(x, y)$ by means of $g(a, b)$ yields the gradient of f_t as a function of the gradient of g. Plugging (14) and (15) into this expression of the gradient of f_t, one gets

$$\partial_x f = 4x + \frac{8}{3}(g - 2a),$$
$$\partial_y f = 2y - 12ty/(g + a).$$

The left hand member of (5) can then be written as

$$(\partial_x f)^2 + (\partial_y f)^2 = 16x^2 + 4y^2 + R,$$

where the remaining term R is

$$R = 8x \frac{8}{3}(g - 2a) + \frac{64}{9}(g - 2a)^2 - 4y\, 12ty/(g + a) + 144t^2y^2/(g + a)^2. \tag{16}$$

Using the expression (9) of H, one can eliminate y from (16) through

$$12ty^2/(g + a)^2 = \frac{16}{9}(g - 2a). \tag{17}$$

Plugging (17) into (16), one gets an expression of R as a function of (g, a, x, t). Finally, $a = x + t$ yields $R = 0$, which proves the claim.

Figure 14 plots the image of the surface S_2 of equation $z = f_2(x, y)$, with front lighting. The image is

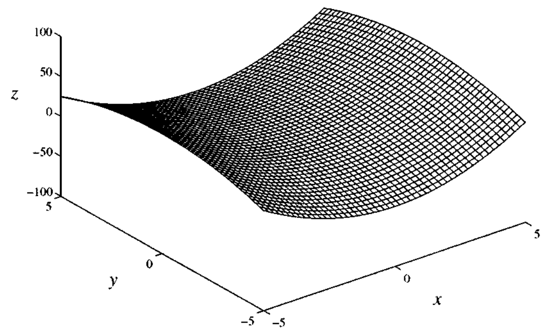

Figure 13. Surface S_H: $z = 2x^2 - y^2$.

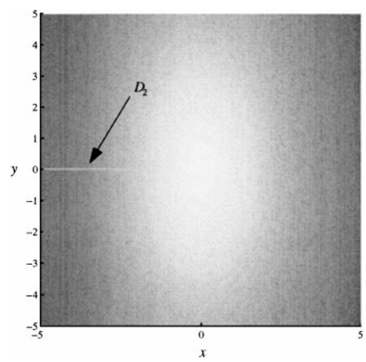

Figure 14. S_2 lit from the direction of the observer.

indeed identical to the one of Fig. 2, with the exception of $D_2\backslash\{P_2\}$, where f_2 is not differentiable.

5. Conclusion

As mentioned in the introduction, an application of Theorem 1 is the perhaps surprising possibility of non-visible smooth deformations. We simulated the deformation of the surface S_0, represented in Fig. 10, into the surface S_t for $t \in [0, 1]$, checking that the image remains indeed identical. Notice that one needs no hypothesis about the behaviour of the surface S with respect to the re-emission of light, i.e., about the function $r((\partial_x f)^2 + (\partial_y f)^2)$ of (2).

On the other hand, if one modifies the direction of the observation or of the lighting, the deformations that we built in Section 4 become visible. Such deformations are invisible only in the context of monovision, as opposed to stereovision which, therefore, seems to prevent efficiently, from the practical point of view, the ambiguity in shape from shading demonstrated in this paper. This shows, if necessary, the superiority of the stereovision methods over the monovision methods for the problems of shape reconstruction.

References

1. A. Blake, A. Zisserman, and G. Knowles, "Surface descriptions from stereo and shading," *Image and Vision Computing*, Vol. 3, No. 4, pp. 183–191, 1985.
2. M.J. Brooks, W. Chojnacki, and R. Kozera, "Shading without shape," *Quarterly of Applied Mathematics*, Vol. L, No. 1, pp. 27–38, 1992.
3. A.R. Bruss, "The Eikonal equation: Some results applicable to computer vision," *Journal of Mathematical Physics*, Vol. 23, No. 5, pp. 890–896, 1982.
4. H. Cartan, *Théorie élémentaire des fonctions analytiques d'une ou plusieurs variables complexes*, pp. 138–139, Hermann, Paris, 1961.
5. B.K.P. Horn, "Obtaining shape from shading information," in *The Psychology of Computer Vision*, Winston: New York, 1975, Ch. 4, pp. 115–155.
6. B.K.P. Horn and M.J. Brooks, "The variational approach to shape from shading," *Computer Vision, Graphics, and Image Processing*, Vol. 33, No. 2, pp. 174–208, 1986.
7. B.K.P. Horn, R.S. Szeliski, and A.L. Yuille, "Impossible shaded images," *IEEE PAMI*, Vol. 15, No. 2, pp. 166–169, 1993.
8. P.L. Lions, E. Rouy, and A. Tourin, "Shape-from-shading, viscosity solutions and edges," *Numer. Math.*, Vol. 64, No. 3, pp. 323–353, 1993.
9. J. Oliensis, "Uniqueness in shape from shading," *International Journal of Computer Vision*, Vol. 6, No. 2, pp. 75–104, 1991.

Jean-Denis Durou received a Ph.D. degree from the Université Paris XI, Orsay, France, in 1993. He is currently working in the field of Computer Vision at the Institut de Recherche en Informatique, Université Paul Sabatier, Toulouse, France.

Didier Piau received a Ph.D. degree from the Université Claude Bernard, Lyon, France, in 1994. He is currently working in the field of Probability Theory at the Laboratoire de Probabilités, Université Claude Bernard, Lyon, France.