Computer Vision and Image Understanding **64**:3 (1996), 390–398. Received September 19, 1995

Regularity properties of distance transformations in image analysis

Christer O. Kiselman

Department of Mathematics, Uppsala University, P. O. Box 480, S-751 06 Uppsala, Sweden

E-mail: cok@math.uu.se

Contents:

Resumo: *Reguleco de distancaj transformoj en la analizo de bildoj*
Distanca transformo donas por ĉiu aro en la bilda ebeno funkcion kiu iel mezuras la distancon al la aro. Multaj malsamaj metodoj por mezuri la distancon eblas, kaj kelkaj havas nedezirindajn ecojn. Montriĝas ke multaj gravaj distancoj povas esti difinitaj per infima kunfaldo. Uzante tiun ĉi ni studas regulecon en la senco de Borgefors kaj starigas kondiĉojn kiuj ekvivalentas al tiu eco.

Abstract: A distance transformation gives for every set in the image plane a function which measures the distance to the set. Several different methods of measuring the distance are possible, and some have undesirable properties. It turns out that many important distances can be defined by infimal convolution. Using this operation we study semiregularity in the sense of Borgefors and give conditions on the distance which are equivalent to this property.

1. Introduction

Distance transformations of digital images are a useful tool in image analysis. A distance transform of a shape is the set of distances from a given pixel to the shape. The distances can be measured in different ways, e.g., by approximating the Euclidean distance in the two-dimensional image, the Euclidean distance between two pixels $x = (x_1, x_2)$ and $y = (y_1, y_2)$ being $\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$. Other distances that have been used are the city-block distance $|x_1 - y_1| + |x_2 - y_2|$ and the chess-board distance $\max(|x_1 - y_1|, |x_2 - y_2|)$.

While the Euclidean distance is easy to visualize geometrically, it has certain drawbacks in this context: in the calculation, we need to keep in memory a vector rather than a scalar at each pixel; we need more operations per pixel; and, perhaps most importantly, the Euclidean distance is more difficult to use for various morphological operations, such as skeletonizing, than for instance the city-block distance; see Borgefors [3]. For a study of the computation of the Euclidean distance transform in any dimension, see Ragnemalm [9].

In the case of the city-block and chess-board distances, one first defines the distances between neighboring pixels; we shall call them, following Starovoitov [13], *prime distances*. Then the distance between any two pixels is defined by following a path and taking as the distance the minimum over all admissible paths of the sum of the prime distances. As an example, for the city-block distance the admissible paths consists of horizontal and vertical moves only, and the prime distance between two pixels which share a side is declared to be one. Thus the distance is calculated successively from neighboring pixels, which is convenient both for sequential and parallel computation.

In this paper we shall give a unified treatment of distances created in this way. It turns out that the metric is conveniently defined from the prime distances by a well-known procedure called infimal convolution over all grid points.

We then consider a regularity property of distance transforms introduced by Borgefors [2] and called semiregularity by her. We give several conditions on a distance function which are necessary and sufficient for semiregularity in two dimensions. Also in higher dimensions a characterization of semiregularity is given.

2. Distances and metrics

Let X be any nonempty set. We shall measure distances between points in X, which amounts to defining a real-valued function on the Cartesian product $X \times X$ of X with itself. Let us agree to call a function $d\colon X \times X \to \mathbf{R}$ a *distance* if d is positive definite:

$$d(x, y) \geqslant 0 \text{ with equality precisely when } x = y, \tag{2.1}$$

and symmetric:

$$d(x, y) = d(y, x) \text{ for all } x, y \in X. \tag{2.2}$$

A distance will be called a *metric* if in addition it satisfies the triangle inequality:

$$d(x, z) \leqslant d(x, y) + d(y, z) \text{ for all } x, y, z \in X. \tag{2.3}$$

All nonempty sets can be equipped with a metric, viz. the discrete metric d_0 defined as

$$d_0(x, x) = 0, \qquad d_0(x, y) = 1 \text{ if } x \neq y. \tag{2.4}$$

We shall use the word *distance* in a noncommital way, thus not implying that it is a metric.

For us the set X will usually be the image plane $\mathbf{Z}^2$ consisting of all points in the plane with integer coordinates (the pixels), or more generally the image space $\mathbf{Z}^n$. Here also a group operation is defined, viz. vector addition: $\mathbf{Z}^2$ is an abelian group. Whenever X is an abelian group it is of particular interest to use translation-invariant distances, i.e., those which satisfy

$$d(x - a, y - a) = d(x, y) \text{ for all } a, x, y \in X. \tag{2.5}$$

The following result is well known and easy to prove; we include it for ease of reference.

Lemma 2.1. *Any translation-invariant distance d on an abelian group X defines a function $g(x) = d(x, 0)$ on X which is positive definite:*

$$g(x) \geqslant 0 \text{ with equality precisely when } x = 0; \tag{2.6}$$

and symmetric:

$$g(-x) = g(x) \text{ for all } x \in X. \tag{2.7}$$

Conversely, a function g which satisfies (2.6) and (2.7) defines a distance $d(x, y) = g(x - y)$.

Lemma 2.2. *Let d be a translation-invariant distance on an abelian group X and g a function on X related to d as in Lemma 2.1. Then d is a metric if and only if g is subadditive:*

$$g(x + y) \leqslant g(x) + g(y) \text{ for all } x, y \in X. \tag{2.8}$$

Proof. If d is a metric, we can write, using the triangle inequality and the translation invariance,

$$g(x + y) = d(x + y, 0) \leqslant d(x + y, y) + d(y, 0) = d(x, 0) + d(y, 0) = g(x) + g(y).$$

Conversely, if g is subadditive,

$$d(x, z) = g(x - z) \leqslant g(x - y) + g(y - z) = d(x, y) + d(y, z),$$

proving the triangle inequality.

A particular kind of distances on an abelian group X are those which are positively homogeneous, i.e., those which satisfy

$$d(mx, my) = md(x, y) \text{ for all } x, y \in X \text{ and all } m \in \mathbf{N}. \tag{2.9}$$

Of course, if g is the function $g(x) = d(x, 0)$ related to a translation-invariant distance, this is equivalent to g being positively homogeneous:

$$g(mx) = mg(x) \text{ for all } x \in X \text{ and all } m \in \mathbf{N}, \tag{2.10}$$

and, in view of the symmetry, also to g being homogeneous:

$$g(mx) = |m|g(x) \text{ for all } x \in X \text{ and all } m \in \mathbf{Z}. \tag{2.11}$$

In these formulas the product mx with $m \in \mathbf{N}$ is interpreted as the sum of m terms all equal to x, and $mx = (-m)(-x)$ when m is a negative integer.

We shall also need the notion of midpoint convexity: a function f on an abelian group X with values in the extended real line $[-\infty, +\infty]$ is said to be *midpoint convex* if it satisfies

$$f(x) \leqslant \tfrac{1}{2} f(x + y) \dotplus \tfrac{1}{2} f(x - y) \text{ for all } x, y \in X. \tag{2.12}$$

Here $\dotplus$ is an extension of addition, called *upper addition* and serving to resolve the problem of conflicting infinities so that $+\infty$ always wins over $-\infty$: one defines

$+\infty \dotplus a = +\infty$ for all elements a of $[-\infty, +\infty]$, and $-\infty \dotplus a = -\infty$ for all $a \in [-\infty, +\infty[$. If the conflict between $+\infty$ and $-\infty$ does not arise, for instance if the function never takes the value $-\infty$, we shall write $+$ instead of $\dotplus$ in (2.12).

It is easy to see that a positively homogeneous function is midpoint convex precisely when it is subadditive, i.e.,

$$f(x+y) \leqslant f(x) \dotplus f(y) \text{ for all } x, y \in X. \tag{2.13}$$

Proposition 2.3. *Let g be a subadditive function on an abelian group X satisfying $0 \leqslant g < +\infty$. Then the following three conditions are equivalent:*
A. g is midpoint convex and $g(0) = 0$;
B. $g(2x) = 2g(x)$ for all $x \in X$;
C. g is positively homogeneous.

Proof. If g is midpoint convex, then taking $y = x$ in (2.12) yields

$$g(x) \leqslant \tfrac{1}{2} g(x+x) + \tfrac{1}{2} g(x-x) = \tfrac{1}{2} g(2x) + g(0),$$

so that $2g(x) \leqslant g(2x)$ if we assume in addition that $g(0) = 0$. The opposite inequality follows from the subadditivity. Thus A implies B.

Now assume that B holds. By subadditivity $g(2x) \leqslant g(x+y) + g(x-y)$. Here the left-hand side is equal to $2g(x)$ if B holds, so g is midpoint convex. Moreover, $g(0) = 2g(0)$, so that $g(0) = 0$, proving A.

Since B is a special case of C, it only remains to be proved that B implies C. For any fixed number $m \in \mathbf{N}$, denote by C(m) the statement that $g(mx) = mg(x)$ for all x. Thus B is the statement C(2), and by repeated application we see that B implies C(2^k) for any $k \in \mathbf{N}$. Next we shall prove that C(m) implies C$(m-1)$ for $m \geqslant 1$. By subadditivity we have

$$g(mx) = g((m-1)x + x) \leqslant g((m-1)x) + g(x) \leqslant (m-1)g(x) + g(x) = mg(x).$$

If C(m) holds, the first and last elements of this inequality are equal, so we must have equality throughout, in particular $g((m-1)x) + g(x) = (m-1)g(x) + g(x)$. If $g(x)$ is finite, which we assume, then it follows that C$(m-1)$ holds. By induction C$(2^k - j)$ holds for all $k, j \in \mathbf{N}$ with $2^k - j \geqslant 0$, which means that C holds.

3. Metrics defined by infimal convolution

Let f, g be two functions defined on an abelian group X with values in the extended real line $[-\infty, +\infty]$. The *infimal convolution* $f \mathbin{\square} g$ of f and g is by definition

$$(f \mathbin{\square} g)(x) = \inf_{y \in X} \big(f(x-y) \dotplus g(y) \big), \quad x \in X. \tag{3.1}$$

Here $\dotplus$, upper addition, is the extension of addition which was defined after (2.12). It amounts to the same to use only those y which satisfy $f(x-y), g(y) < +\infty$ in (3.1). For a survey of the properties of this operation, see Moreau [8], Rockafellar [10], or Strömberg [14].

In the definition of an infimal convolution the infimum operator acts over an infinite set of points, and therefore sometimes cannot be computed in finitely many steps. However, there are many situations where the infimum is in fact a minimum over a finite set. One such case is when f is bounded from below and g is coercive in the strong sense that all sublevel sets $\{y;\ g(y) \leqslant a\}$, $a \in \mathbf{R}$, are finite. Then in particular the sublevel set $\{y;\ g(y) \leqslant (f \mathbin{\square} g)(x) + 1 - \inf f\}$ is finite, and it is enough to search for a minimizing y in that set. Even simpler is the case when g is less than $+\infty$ in a finite set P only. Then any infimal convolution with g is equal to the minimum

$$(f \mathbin{\square} g)(x) = \min_{y \in P} \big(f(x-y) \dot{+} g(y)\big), \quad x \in X. \tag{3.2}$$

This is indeed the case for the distances we shall consider: here P is a small set around the origin where the prime distances are defined.

We have seen that subadditive functions are important when it comes to defining metrics (Lemma 2.2). Therefore it is of interest to know that subadditivity can be characterized using infimal convolution:

Lemma 3.1. *A function f on an abelian group is subadditive (in the sense of (2.13)) if and only if it satisfies the inequality $f \mathbin{\square} f \geqslant f$. If $f(0) = 0$, this is equivalent to the equation $f \mathbin{\square} f = f$.*

Proof. If f is subadditive we have $f(x-y) \dot{+} f(y) \geqslant f(x)$, so taking the infimum over y gives $(f \mathbin{\square} f)(x) \geqslant f(x)$. Conversely, $f(x) \dot{+} f(y) \geqslant (f \mathbin{\square} f)(x+y)$ for all x, y, so $f \mathbin{\square} f \geqslant f$ implies subadditivity. Finally, we always have $(f \mathbin{\square} f)(x) \leqslant f(x) \dot{+} f(0)$, so if $f(0) = 0$ it follows that $f \mathbin{\square} f \leqslant f$.

Infimal convolution is a commutative and associative operation on functions, so we can write iterated convolutions as $f \mathbin{\square} g \mathbin{\square} h$ without using parentheses. A k-fold convolution can be defined by

$$(f_1 \mathbin{\square} \cdots \mathbin{\square} f_k)(x) = \inf \sum_{i=1}^{k} f_i(x^i), \tag{3.3}$$

where the infimum is over all choices of elements $x^i \in X$ such that $x^1 + \cdots + x^k = x$, and with the understanding that the sum receives the value $+\infty$ as soon as one of the terms has that value, even in the presence of a value $-\infty$. In (3.3) it is natural to think of a path leading from 0 to x consisting of segments $[0, x^1]$, $[x^1, x^1 + x^2]$,..., $[x^1 + \cdots + x^{k-1}, x]$; if $X = \mathbf{Z}^2$ this path can be realized in $\mathbf{R}^2$.

Denoting by $\operatorname{dom} f = \{x \in X;\ f(x) < +\infty\}$ the *effective domain* of f, i.e., the set where f does not take the value $+\infty$, we see that

$$\operatorname{dom}(f_1 \mathbin{\square} \cdots \mathbin{\square} f_k) = \operatorname{dom} f_1 + \cdots + \operatorname{dom} f_k. \tag{3.4}$$

If A is a subset of an abelian group X, we shall write $\mathbf{N} \cdot A$ for the semigroup generated by A:

$$\mathbf{N} \cdot A = \left\{\textstyle\sum m_i a_i;\ m_i \in \mathbf{N}, a_i \in A\right\}, \tag{3.5}$$

where all but finitely many of the m_i are zero. Similarly, we shall write $\mathbf{Z} \cdot A$ for the group generated by A:

$$\mathbf{Z} \cdot A = \left\{\textstyle\sum m_i a_i;\ m_i \in \mathbf{Z}, a_i \in A\right\}. \tag{3.6}$$

If A is symmetric, $A = -A$, then of course $\mathbf{Z} \cdot A = \mathbf{N} \cdot A$.

It seems plausible that if a repeated convolution $G \,\square\, G \,\square\, \cdots \,\square\, G$ has a limit g as the number of factors tends to infinity, then this limit will satisfy the equation $g \,\square\, g = g$. Indeed this is the case under very general hypotheses:

Theorem 3.2. *Let $G\colon X \to [0, +\infty]$ be a function on an abelian group X satisfying $G(0) = 0$. Define a sequence of functions $(G_j)_{j=1}^{\infty}$ by putting $G_1 = G$, $G_j = G_{j-1} \,\square\, G$, $j = 2, 3, ...$, in other words, G_j is the infimal convolution of j factors all equal to G. Then the sequence (G_j) is decreasing and its limit $\lim G_j = g \geqslant 0$ is subadditive. Moreover $\operatorname{dom} g = \mathbf{N} \cdot \operatorname{dom} G$, i.e., g is finite precisely in the semigroup generated by $\operatorname{dom} G$.*

Remark. It is easy to prove that g is the largest subadditive minorant of G.

Proof. That the sequence is decreasing is obvious if we take $y = 0$ in the definition of G_{j+1}:

$$G_{j+1}(x) = \inf_y \bigl(G_j(x-y) + G(y)\bigr) \leqslant G_j(x) + G(0) = G_j(x).$$

Next we shall prove that $g(x+y) \leqslant g(x) + g(y)$. If one of $g(x), g(y)$ is equal to $+\infty$ there is nothing to prove, so let x, y be given with $g(x), g(y) < +\infty$ and fix a positive number ε. Then there exist numbers j, k such that $G_j(x) \leqslant g(x) + \varepsilon$ and $G_k(y) \leqslant g(y) + \varepsilon$. By associativity $G_{j+k} = G_j \,\square\, G_k$, so we get

$$g(x+y) \leqslant G_{j+k}(x+y) \leqslant G_j(x) + G_k(y) \leqslant g(x) + g(y) + 2\varepsilon.$$

Since ε is arbitrary, the inequality $g(x+y) \leqslant g(x) + g(y)$ follows. Finally, the statement about $\operatorname{dom} g$ is an easy consequence of (3.4).

Theorem 3.3. *With G as in Theorem 3.2, assume in addition that there is a metric d_1 on X such that $G(x) \geqslant d_1(x, 0)$ for all $x \in X$. Then the limit g of the sequence G_j also satisfies this inequality, $g(x) \geqslant d_1(x, 0)$, so that it is positive definite. If G is symmetric, g is also symmetric and defines a metric $d(x, y) = g(x-y) \geqslant d_1(x, y)$ on the subgroup $\mathbf{Z} \cdot P = \mathbf{N} \cdot P$ of X generated by $P = \operatorname{dom} G$.*

When applying this theorem we could for instance let d_1 be εd_0, where ε is a small positive number and d_0 is the discrete metric defined by (2.4). In $\mathbf{Z}^n$ we can also use $d_1(x, y) = \varepsilon \|x - y\|$ for any norm on $\mathbf{R}^n$.

We call G a *prime distance function.* This is the term used by Starovoitov [13]; Borgefors [2] calls $G(x-y)$ the *local distances.*

Proof. Define $H(x) = d_1(x, 0)$ and let H_j be the infimal convolution of j factors equal to H. From Lemmas 2.2 and 3.1 it follows that $H \,\square\, H = H$ and so all $H_j = H$. Therefore $G \geqslant H$ implies $G_j \geqslant H$ and also the limit g must satisfy $g \geqslant H$. This proves the theorem.

Corollary 3.4. *Let P be a finite set in an abelian group X containing the origin and let G be a function on X with $G(0) = 0$, taking the value $+\infty$ outside P and finite positive values at all points in $P \smallsetminus \{0\}$. Then $g = \lim G_j$ is a positive definite subadditive function. If P is symmetric and $G(-x) = G(x)$, then g defines a metric on the subgroup $\mathbf{Z} \cdot P = \mathbf{N} \cdot P$ of X generated by P.*

Proof. Since P is finite, there is a positive number ε such that $G(x) \geqslant \varepsilon$ for all $x \in P$ except $x = 0$. Thus $G(x) \geqslant \varepsilon d_0(x, 0)$, where d_0 is the discrete metric defined by (2.4). We can now apply the theorem.

Example 3.5. It is by no means necessary that g be positively homogeneous. In fact, we can let $P = \{0, \pm 1, \pm 2\} \subset \mathbf{Z}$ and define $G(\pm 1) = a$, $G(\pm 2) = b$, where a and b are arbitrary positive numbers. If $b \geqslant 2a$, then $g(x) = a|x|$ for all $x \in \mathbf{Z}$, but if $b < 2a$, then $g(x) = \frac{1}{2}b|x|$ when x is even, $x = 2k$, $k \in \mathbf{Z}$, whereas $g(x) = b|k| + a > \frac{1}{2}b|x|$ when x is odd, $x = \pm(2k+1)$, $k \in \mathbf{N}$. But nevertheless g is subadditive.

Example 3.6. A more interesting example is perhaps this in two dimensions. Let $P = \{x \in \mathbf{Z}^2;\ |x_j| \leqslant 1\}$, and define the prime distances as $G(\pm 1, 0) = G(0, \pm 1) = a > 0$, $G(\pm 1, \pm 1) = b > 0$. Then if $b \geqslant a$ we get $g(x_1, 0) = a|x_1|$. But if $b < a$, then $g(2, 0) = 2b < 2a$, so that $g(2, 0) < 2g(1, 0) = 2a$. In fact, by the definition of infimal convolution, $g(2, 0) \leqslant G_2(2, 0) \leqslant G(1, 1) + G(1, -1) = b + b$. On the other hand, it is not difficult to see that for any k, $G_k(2, 0) \geqslant 2b$, so that actually $g(2, 0) = 2b$. This is because if we take $k \geqslant 2$ nonzero steps to go from the origin to $(2, 0)$, the distance assigned to the path is at least $G(x^1) + \cdots + G(x^k) \geqslant kb$.

Example 3.7. Several metrics on $\mathbf{Z}^2$ have been studied. When presenting the generating function G defining the prime distances it shall be understood in the sequel that G is invariant under permutation and reflection of the coordinates. Therefore it is enough to define $G(x)$ for $0 \leqslant x_2 \leqslant x_1$. Also it is understood that $G(0) = 0$ in all cases, and that $G(x) = +\infty$ when not mentioned. Consider first $P = \{x \in \mathbf{Z}^2;\ \sum |x_j| \leqslant 1\}$ and $G(1, 0) = 1$. Then the corresponding metric is the city-block metric, introduced and studied by Rosenfeld & Pfaltz [11]. If instead we let $P = \{x \in \mathbf{Z}^2;\ |x_j| \leqslant 1\}$ and $G(1, 0) = G(1, 1) = 1$, then the metric is the chess-board metric, introduced by Rosenfeld & Pfaltz [12]. Some other metrics that have been studied are modifications of this; to define them, put $G(1, 0) = a$ and $G(1, 1) = b$. Then the choices $(a, b) = (1, \sqrt{2})$ (Montanari [7]); $(a, b) = (2, 3)$ (Hilditch & Rutovitz [6]); and $(a, b) = (3, 4)$ (Borgefors [1]) have all been studied. Next we can increase the size of the neighborhood where prime distances are defined to include the knight's move $(2, 1)$ as an element of P. The distance defined by this move *only* has been studied by Das & Chatterji [5]. It seems more natural, however, to allow also $(1, 0)$ and $(1, 1)$ in P. Then a very good choice under certain criteria is $G(1, 0) = 5$, $G(1, 1) = 7$, and $G(2, 1) = 11$ (the 5-7-11 weighted distance). This distance was proposed and studied by Borgefors [2].

Example 3.8. We always have $g \leqslant G$, and it may happen that $g(x) < G(x)$ for some pixel $x \in P$. Let for instance $G(1, 0) = a$, $G(2, 1) = c$, and extend G by reflection and permutation of the coordinates. Then

$$g(1, 0) \leqslant G_3(1, 0) \leqslant G(2, 1) + G(1, -2) + G(-2, 1) = 3c,$$

so if $3c < a$ we get $g(1, 0) \leqslant 3c < a = G(1, 0)$. This is undesirable, because we expect the prime distance originally defined between the origin and $(1, 0) \in P$ to survive and to be equal to the distance defined by the minimum over all paths. It is therefore natural to require that $g = G$ everywhere in P.

Remark. Let us say that a metric $d(x, y) = g(x - y)$ is *finitely generated* if it is constructed as in Corollary 3.4. It is easy to prove that the Euclidean metric $d(x, y) = \sqrt{\sum (x_j - y_j)^2}$ on $\mathbf{Z}^n$ is finitely generated if and only if $n \leqslant 1$.

Since $g(x)$ is the limit of an infinite sequence $G_j(x)$, it is reassuring to know that this sequence is in fact stationary in the cases of interest here. It is easy to explicitly give an index j such that $G_j(x)$ is equal to the limit $g(x)$:

Proposition 3.9. *Let G be as in Corollary 3.4. Then the sequence (G_j) is pointwise stationary, i.e., for every $x \in X$ there is an index j_x such that $G_j(x) = g(x)$ for all $j \geqslant j_x$.*

Proof. If $g(x) = +\infty$, then of course $G_j(x) = +\infty$ for all j. Consider a point such that $g(x) < \infty$. Then there is an index m such that $G_m(x) < +\infty$. For every j there are elements $y^1, y^2, ..., y^j$ in P such that $x = y^1 + \cdots + y^j$ and $G_j(x) = G(y^1) + \cdots + G(y^j)$ (cf. (3.2) and (3.3)). Now by hypothesis $G(y) \geqslant \varepsilon > 0$ for all $y \in P$ except $y = 0$, so that $G_j(x) \geqslant j_1\varepsilon$, where j_1 is the number of indices i such that $y^i \neq 0$. If $j \geqslant m$ we conclude that $G_m(x) \geqslant G_j(x) \geqslant j_1\varepsilon$, so that $j_1 \leqslant G_m(x)/\varepsilon = C$. Therefore the number of nonzero terms in the representation of $G_j(x)$ is never larger than C. Now $G_j(x) = G_{j_1}(x)$, so the fact that j_1 is bounded means that the sequence is stationary.

4. Regularity properties

Let G be a function as in Theorem 3.2 and $g = \lim G_j$ the limit of the sequence (G_j) of infimal convolutions of G. We have seen that it may happen that $g < G$ at some point in P (see Example 3.8), but it is not enough to rule out that behavior. For every multiple mx of a point $x \in P$ we have a representation $mx = x + \cdots + x$ (m terms), so that $g(mx) \leqslant G_m(mx) \leqslant mG(x)$. This means that we have a path from 0 to mx consisting of m steps equal to x. This is so to speak the most natural path. But usually there exist also other paths from 0 to mx, in fact infinitely many, and it is not desirable that they give a smaller value of the distance. This is the reason behind the introduction of the notion of semiregularity.

We shall say following Borgefors [3, Definition 2] that a prime distance function G is *semiregular* if $g(mx) = mG(x)$ for all $x \in P$ and all $m \in \mathbf{N}$. (Note that this is a property of G and not of g, since G is not uniquely determined from g. In most cases of interest, however, $g = G$ in P, and then we can say that the metric itself is semiregular, viz. when it satisfies $g(mx) = mg(x)$ for all $x \in P$ and all $m \in \mathbf{N}$.) Now $g(mx)$ is the infimum of all $G_j(mx)$, $j \in \mathbf{N}$, and each $G_j(mx)$ is the infimum of all sums $G(y^1) + \cdots + G(y^j)$ where the $y^i \in P$ and $mx = y^1 + \cdots + y^j$, so semiregularity means that for any representation $mx = y^1 + \cdots + y^j$ with $x, y^i \in P$ we have $mG(x) \leqslant G(y^1) + \cdots + G(y^j)$. Slightly more generally we shall say that an arbitrary function $G\colon X \to [-\infty, +\infty]$ is *semiregular* if

$$G(x) < +\infty, mx = y^1 + \cdots + y^j \text{ implies } mG(x) \leqslant G(y^1) \dotplus \cdots \dotplus G(y^j). \tag{4.1}$$

Borgefors [3] also introduced the notion of regularity: G is said to be *regular* if for any point mx with $x \in P$ and $m \in \mathbf{N}$ and any representation $mx = y^1 + \cdots + y^j$ with $y^i \in P$ but not all equal to x or 0 we have a strict inequality $G(y^1) + \cdots + G(y^j) > mG(x)$.

Thus regularity means that the straight line is the unique minimal path from 0 to mx, whereas semiregularity means that the straight line from 0 to mx is minimal, but not necessarily the only minimal path.

Summing up, we conclude that the prime distance function G always defines a metric via the limit $g = \lim G_j$, but that this metric agrees with the one we would like to have on the rays $\mathbf{N}y$, $y \in P$, if and only if G is semiregular.

Example 4.1. As an example, consider the distance which is like the chess-board distance but assigns the value a to the points $(\pm 1, 0)$ and $(0, \pm 1)$ and the value b to the four points $(\pm 1, \pm 1)$; see Example 3.7. It is regular if and only if $a < b < 2a$, and semiregular if and only if $a \leqslant b \leqslant 2a$; see Borgefors [3]. For instance, there are many admissible paths from $(0,0)$ to $(2,0)$, and the path $(1,0)+(1,0)$ is assigned the value $2a$, the path $(1,1)+(1,-1)$ the value $2b$; therefore semiregularity implies $2a \leqslant 2b$, regularity $2a < 2b$.

There is a sufficient condition for semiregularity which we note for later reference:

Proposition 4.2. *Let $G: X \to [-\infty, +\infty]$ be a function on an abelian group X. Assume that there exists a function f which is positively homogeneous, midpoint convex and agrees with G wherever G is less than $+\infty$. Then G is semiregular.*

Proof. Let $mx = y^1 + \cdots + y^j$. We shall prove that $mG(x) \leqslant G(y^1) \dotplus \cdots \dotplus G(y^j)$ when $x \in \operatorname{dom} G$. If one of the $G(y^i) = +\infty$, this inequality certainly holds; on the other hand, if $y^i \in \operatorname{dom} G$ for all i, we know that $G(x) = f(x)$ and $G(y^i) = f(y^i)$, so that the inequality follows from the subadditivity of f:

$$mG(x) = mf(x) = f(mx) \leqslant f(y^1) \dotplus \cdots \dotplus f(y^j) = G(y^1) \dotplus \cdots \dotplus G(y^j).$$

5. Semiregularity of distances in two dimensions

The results up till now have a sense for all abelian groups X. In this section we let X be the image plane $\mathbf{Z}^2$; we can then embed it into the Euclidean plane $\mathbf{R}^2$ and use also functions defined there.

A very natural idea to construct a distance in $\mathbf{Z}^2$ from a function G defining prime distances is the following, used by Montanari [7: Theorem 1]. We first define it on all rays $\mathbf{R}^+p^i$ defined by P, i.e., we define $f(sp^i) = sG(p^i)$ for $s \in \mathbf{R}^+$ and p^i an element of P. This makes sense if two different rays $\mathbf{R}^+p^i$ and $\mathbf{R}^+p^j$ intersect only at the origin, so we assume this to be true. Here we denote by $\mathbf{R}^+$ the set of all nonnegative real numbers, so that $\mathbf{R}^+p$ is the ray from the origin through p:

$$\mathbf{R}^+p = \{tp;\ t \in \mathbf{R},\ t \geqslant 0\}.$$

We then extend f to all of $\mathbf{R}^2$ so that it becomes linear in each sector defined by two neighboring vectors in P. To make this precise we let the nonzero elements of P be $p^1, ..., p^k$ enumerated in the counterclockwise direction, so that p^1 and p^2 define a sector free from elements of P, and so on until the sector defined by p^k and p^1. Then in the sector defined by p^i and p^{i+1} we define

$$f(sp^i + tp^{i+1}) = sG(p^i) + tG(p^{i+1}), \quad s, t \in \mathbf{R}^+. \tag{5.1}$$

Here of course p^{i+1} shall be understood as p^1 if $i = k$.

The function f will then actually be piecewise linear on $\mathbf{R}^2$, and defines a distance there as well as on $\mathbf{Z}^2$. Is it also convex? Is the distance a metric? It turns out that the answers to these questions are intimately connected to the notion of semiregularity of the prime distance function in the sense of Borgefors.

Theorem 5.1. *Let a finite symmetric set P in $\mathbf{Z}^2$ be given, $P = \{p^0, p^1, ..., p^k\}$, where $p^0 = 0$. Assume that two rays $\mathbf{R}^+ p^i$ and $\mathbf{R}^+ p^j$, $i \neq j$, intersect only at the origin and that P contains two linearly independent vectors. Let also a symmetric function G be given with finite positive values in $P \smallsetminus \{0\}$, $G(0) = 0$, and the value $+\infty$ outside P. Then define f in $\mathbf{R}^2$ to be equal to G in P, to be positively homogeneous and piecewise linear in each sector which does not contain any point from P in its interior. Explicitly, this means that we define f by (5.1) above. The following five conditions are equivalent:*

A. The prime distance function G is semiregular in the sense of Borgefors;
B. The function f is (midpoint) convex in $\mathbf{R}^2$;
C. The restriction of f to $\mathbf{Z}^2$ is midpoint convex;
D. The distance $d_f(x, y) = f(x - y)$ is a metric on $\mathbf{R}^2$;
E. The distance $d_f(x, y) = f(x - y)$ is a metric on $\mathbf{Z}^2$.

Remark. It is well known and not difficult to prove that a midpoint convex function on a real vector space which has some kind of boundedness must be convex; it satisfies $f((1 - \lambda)x + \lambda y) \leqslant (1 - \lambda)f(x) + \lambda f(y)$ for all $\lambda \in [0, 1]$. (Using the axiom of choice one can construct midpoint convex functions which are not convex, but they must be unbounded both from below and from above near every point.) In our case the function is continuous, so midpoint convexity is equivalent to convexity. We also note that $f \geqslant g$ on $\mathbf{N}p^i + \mathbf{N}p^{i+1}$ if g is the function constructed from G as in Corollary 3.4. Indeed,

$$f(kp^i + mp^{i+1}) = kG(p^i) + mG(p^{i+1}) \geqslant G_{k+m}(kp^i + mp^{i+1}) \geqslant g(kp^i + mp^{i+1}).$$

It may happen that $g(x) > f(x)$ for certain pixels x.

Proof. To start at the easy end, let us first point out that B and C are equivalent. In fact B implies C by taking the restriction from $\mathbf{R}^2$ to $\mathbf{Z}^2$. If C holds, it follows from the homogeneity that $f(x) \leqslant \frac{1}{2}f(x+y) + \frac{1}{2}f(x-y)$ for all vectors with rational components, and then for all vectors by continuity. The proof that D and E are equivalent is of course similar.

Next we shall first prove the equivalence of B and D (and of C and E; the proof is the same):

Lemma 5.2. *Let f be a real-valued function defined on an abelian group X such that*

$$f(-x) = f(x) \geqslant 0 \text{ with equality only for } x = 0; \tag{5.2}$$

and

$$f(2x) = 2f(x) \text{ for all } x \in X. \tag{5.3}$$

Define a distance on X by $d(x, y) = f(x - y)$. Then d is a metric if and only if f is midpoint convex.

Proof. The properties (2.1) and (2.2) of a metric being obviously fullfilled, the only question can be whether d satisfies the triangle inequality (2.3). If f is midpoint convex we get

$$\begin{aligned} d(x, z) = f(x - z) &\leqslant \tfrac{1}{2}f(2x - 2y) + \tfrac{1}{2}f(2y - 2z)) \\ &= f(x - y) + f(y - z) = d(x, y) + d(y, z), \end{aligned}$$

so the triangle inequality is true.

Conversely, suppose now that the triangle inequality holds. Then f is midpoint convex:

$$2f(x) = f(2x) = d(2x, 0) \leqslant d(2x, x-y) + d(x-y, 0) = f(x+y) + f(x-y).$$

This concludes the proof of the lemma.

Proof of Theorem 5.1, continued. Having established the equivalence of B, C, D, and E, we shall now see that B implies A. Indeed, if f is convex, then G is semiregular by Proposition 4.2. (In this case we have $g \geqslant f$ everywhere in $\mathbf{R}^2$.)

Finally we shall prove that A implies B. Thus assume that the prime distance function G is semiregular. We have to prove that f is convex, but since it is piecewise linear in the sectors defined by the vectors in P, it is enough to prove that the linear interpolation $f_{i-1,i+1}$ of f between the rays $\mathbf{R}^+p^{i-1}$ and $\mathbf{R}^+p^{i+1}$ lies above f on the ray $\mathbf{R}^+p^i$. The function $f_{i-1,i+1}$ is given by

$$f_{i-1,i+1}(sp^{i-1} + tp^{i+1}) = sG(p^{i-1}) + tG(p^{i+1}), \quad s, t \in \mathbf{R}^+,$$

cf. (5.1). Now the value of $f_{i-1,i+1}$ at a point $x = sp^{i-1} + tp^{i+1}$ in the sector defined by p^{i-1} and p^{i+1} with $s, t \in \mathbf{N}$ is precisely the length assigned to the path from 0 to $sp^{i-1} + tp^{i+1}$ consisting of the segment from 0 to sp^{i-1} followed by the segment from that point to $x = sp^{i-1} + tp^{i+1}$. Suppose now that the latter point is on the ray $\mathbf{R}^+p^i$, thus $sp^{i-1} + tp^{i+1} = rp^i$ for some $r \in \mathbf{R}^+$. By semiregularity, the value of that length is not smaller than the value of f at rp^i, assuming s, t and r to be integers. This means that $f \leqslant f_{i-1,i+1}$ at the point rp^i. In general, if $s, t \in \mathbf{N}$, r will be a rational number. Therefore r will be an integer if we choose s and t as multiples of some integer. In view of the positive homogeneity of f and $f_{i-1,i+1}$ we must then have $f \leqslant f_{i-1,i+1}$ on the whole ray $\mathbf{R}^+p^i$, which, as we remarked, means that f is convex.

6. Semiregularity of distances in any dimension

Distance transformations are of interest also in higher dimensions, cf. Borgefors [4]. We shall therefore take at look at distances defined in $\mathbf{Z}^3$, which we identify with the space of voxels, and more generally in $\mathbf{Z}^n$ for any n. We embed that group into the vector space $\mathbf{R}^n$ and find that it is easy to formulate a converse to Proposition 4.2:

Theorem 6.1. *Let $G\colon \mathbf{R}^n \to [0, +\infty]$ be any given function. If there exists a convex positively homogeneous function which agrees with G in $\operatorname{dom} G$, then G is semiregular. If all coordinates of all points in $\operatorname{dom} G$ are rational, then the converse holds.*

Proof. The first part is a special case of Proposition 4.2. For the converse, let F_G denote the largest minorant of G which is convex and positively homogeneous. We can describe it as

$$F_G(x) = \inf \sum \lambda_i G(y^i), \quad x \in \mathbf{R}^n,$$

where the infimum is taken over all points y^i and all positive numbers λ_i such that $\sum \lambda_i y^i = x$ (a finite linear combination). If there is some convex and positively homogeneous function which agrees with G in $\operatorname{dom} G$, then F_G must have this property. We thus want to prove that $F_G \geqslant G$ in $\operatorname{dom} G$, which amounts to

$$G(x) \leqslant \sum \lambda_i G(y^i) \tag{6.1}$$

for all representations of x as a finite linear combination $x = \sum \lambda_i y^i \in \operatorname{dom} G$, $\lambda_i > 0$. We see that semiregularity is a special case of this, viz. $\lambda_i = 1/m$. Since we can lump the y^i together if some of them are equal, semiregularity implies that (6.1) holds if all the λ_i are positive and rational.

Now if all components of x and the y^i are rational, then by classical linear algebra, to every representation $x = \sum \lambda_i y^i$ with λ_i real there is a representation $x = \sum \mu_i y^i$ with μ_i rational and arbitrarily close to λ_i, so also $\sum \mu_i G(y^i)$ is as close to $\sum \lambda_i G(y^i)$ as we wish. It follows that if (6.1) holds for λ_i rational and positive, then it also holds for λ_i real and positive. This proves the theorem.

In Theorem 5.1, the piecewise linear function f is convex precisely when G admits a convex and positively homogeneous interpolating function. But in the construction of f we used a circular ordering of the elements of P, and this is no longer possible in three or more dimensions: the relation between P and the function f has to be described somewhat differently.

Consider a positively homogeneous function $f\colon \mathbf{R}^n \to [0, +\infty]$ which is piecewise linear, i.e., linear in each of a finite number of cones. This means that there exists a finite set P in $\mathbf{R}^n$, $P = \{p^0, p^1, ..., p^k\}$, $p^0 = 0$, which defines the cones in the following way. For every subset I of the set of indices $\{1, ..., k\}$ let $C(I)$ denote the convex cone spanned by the p^i with $i \in I$:

$$C(I) = \left\{\textstyle\sum_{i\in I} s_i p^i \in \mathbf{R}^n;\ s_i \geqslant 0\right\}.$$

Each set $C(I)$ is a closed cone which may or may not have interior points. Then for any $x \in \mathbf{R}^n$ there is a set I_x of at most n indices such that f is linear in the cone $C(I_x)$. This means that

$$y = \sum_{i\in I_x} t_i p^i, t_i \geqslant 0, \text{ implies } f(y) = \sum_{i\in I_x} t_i f(p^i). \tag{6.2}$$

If we let x vary, we see that the union of the (finitely many) $C(I_x)$ cover $\mathbf{R}^n$, so that those of them which do possess interior points also cover $\mathbf{R}^n$.

A difficulty in dimensions three and higher is that the restriction of f to P does not define f. (To illustrate this, let P be the set consisting of the four points $(0, \pm 1, 1), (\pm 1, 0, 1)$ in $\mathbf{R}^3$. In the convex cone spanned by these points we can consider two piecewise linear functions $f_1(x) = |x_3| + |x_1|$ and $f_2(x) = 2|x_3| - |x_2|$; they have the same restriction to P and none is a more canonical extension than the other.) Also of course the function f does not determine the set P uniquely; we may add points at will. Such redundant points will change the restriction G of f to P as well as the metric g defined by G. Also, we can multiply the elements of P by different positive scalars without affecting its relation to f. This will change both G and the metric g. We shall therefore have to assume that both f and P are given. The result below holds for all G obtained in this way.

Theorem 6.2. *Let $f\colon \mathbf{R}^n \to [0, +\infty]$ be positively homogeneous and piecewise linear, and let P be a finite set in $\mathbf{R}^n$, $P = \{p^0, p^1, ..., p^k\}$, $p^0 = 0$, which describes f as in (6.2). Let us assume (in order to get a distance) that f and P are symmetric, i.e., $f(-x) = f(x)$ and $-P = P$. Define G to be equal to f in P and equal to $+\infty$ outside P. Consider the following five conditions:*

A. The prime distance function G is semiregular in the sense of Borgefors;
B. The function f is (midpoint) convex in $\mathbf{R}^n$;
C. The restriction of f to $\mathbf{Z}^n$ is midpoint convex;
D. The distance $d_f(x,y) = f(x-y)$ is a metric on $\mathbf{R}^n$;
E. The distance $d_f(x,y) = f(x-y)$ is a metric on $\mathbf{Z}^n$.

Then B–E are all equivalent. They imply A but not conversely for $n \geqslant 3$.

Concerning property A we should remark that G defines a metric via infimal convolution on the subgroup $\mathbf{Z} \cdot P$ generated by P; this subgroup may or may not be a subgroup of $\mathbf{Z}^n$. The property A depends on the choice of P and is stronger the larger P is, but it is desirable to choose a minimal P.

Proof of Theorem 6.2. All implications in the proof of Theorem 5.1 except A implies B are valid without change. This proves the theorem except for the last remark, which will follow from the next example.

Example 6.3. Let $f(x) = a \max |x_j| + b \min |x_j|$, $x \in \mathbf{R}^n$, where a and b are positive numbers. This function is positively homogeneous and piecewise linear. Let P be the set consisting of the 3^n points in $\mathbf{Z}^n$ with $|x_j| \leqslant 1$. This set is sufficient to describe f using (6.2). Let G be equal to f in P and to $+\infty$ outside. If $n = 2$, $f(x) = b(|x_1| + |x_2|) + (a-b) \max |x_j|$ is convex if and only if $b \leqslant a$, and the prime distance function G is semiregular simultaneously—in agreement with Theorem 5.1. If $n \geqslant 3$, f is never convex. In fact, $f(x_1, 1, ..., 1, 2) = 2a + b \min(x_1, 1)$ when $0 \leqslant x_1 \leqslant 2$, and this is not a convex function of x_1. Then G is semiregular if and only if $(n-1)b \leqslant a$. In fact, if there is a function f_1 which agrees with f on P and is convex and positively homogeneous, then

$$a + b = f_1(1, ..., 1) = \frac{n}{n-1} f_1((n-1)/n, ..., (n-1)/n) \leqslant \frac{n}{n-1} a,$$

since a is the value of f at the n points with one zero and $n-1$ ones, and since $((n-1)/n, ..., (n-1)/n)$ is the barycenter of these points. Hence $a + b \leqslant na/(n-1)$ is a necessary condition for the existence of such a function f_1. It is also sufficient, since we may define

$$f_1(x) = \max\left(a|x_k|, \frac{a+b}{n} \sum |x_j|\right), \quad x \in \mathbf{R}^n,$$

which always gives the wanted value $a+b$ at the point $(1, ..., 1)$, and the wanted value a at the point $(0, 1, ..., 1)$ if $(a+b)(n-1)/n \leqslant a$, i.e., if $(n-1)b \leqslant a$. We can conclude in view of Theorem 6.1 that G is semiregular if and only if $(n-1)b \leqslant a$, but, as already remarked, f is never convex when $n \geqslant 3$.

7. Conclusion

In this paper we have shown that the operation of infimal convolution offers a convenient formalism for the construction of distances in the image plane: most of the distances used in image analysis can be obtained as limits of a sequence of convolution products of a prime distance function defining the distances between neighboring pixels (Corollary 3.4). The triangle inequality for a translation-invariant distance is easily expressed as an infimal convolution equation (Lemmas 2.2 and 3.1). This is true in all dimensions.

It has also been shown that, in two dimensions, semiregularity in the sense of Borgefors [2] of these distances can be expressed in terms of midpoint convexity of an associated function; alternatively, in terms of the triangle inequality of the distances (Theorem 5.1). In all dimensions, semiregularity is equivalent to the existence of a convex interpolating function (Theorem 6.1).

Acknowledgment

The origin of this paper lies in a talk on mathematical problems in image analysis given by Gunilla Borgefors at Uppsala University on March 6, 1995. I am very grateful for her help and encouragement concerning these problems, in particular concerning the regularity conditions considered here, which were presented in her talk.

References

1. G. Borgefors, Distance transformations in arbitrary dimensions, *Comput. Vision Graphics Image Process.* **27**, 1984, 321–345.
2. —, Distance transformations in digital images, *Comput. Vision Graphics Image Process.* **34**, 1986, 344–371.
3. —, Applications using distance transforms, in *Aspects of Visual Form Processing,* pp. 83–108, (C. Arcelli & S. di Baja, Eds.), World Scientific, Singapore, 1994.
4. —, On digital distance transforms in three dimensions. To appear in *Computer Vision and Image Understanding.*
5. P. P. Das & B. M. Chatterji, Knight's distance in digital geometry, *Pattern Recognition Letters* **7**, 1988, 215–226.
6. J. Hilditch & D. Rutovitz, Chromosome recognition, *Annals of the New York Academy of Sciences* **157**, 1969, 339–364.
7. U. Montanari, A method for obtaining skeletons using a quasi-Euclidean distance, *J. Assoc. Comput. Mach.* **15**, 1968, 600-624.
8. J.-J. Moreau, Inf-convolution, sous-additivité, convexité des fonctions numériques, *J. Math. Pures et Appl.* **49**, 1970, 109–154.
9. I. Ragnemalm, The Euclidean distance transform in arbitrary dimensions, *Pattern Recognition Letters* **14**, 1993, 883–888.
10. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.
11. A. Rosenfeld & J. Pfaltz, Sequential operations in digital picture processing, *Journal of the ACM* **13**, 1966, 471–494.
12. —, Distance functions in digital pictures, *Pattern Recognit.* **1**, 1968, 33–61.
13. V. Starovoitov, Towards a distance transform generalization, in *Proceedings of the 9th Scandinavian Conference on Image Analysis,* pp. 499-506, (G. Borgefors, Ed.), Uppsala, 1995.
14. T. Strömberg, The operation of infimal convolution. *Dissertationes Math.* **352** (1996), 58 pp.