Math 574

Divided Cells

Spring 2004

1. Reference and definition

The formulation of the "divided cell" algorithm as an approach to the study of the values of a function $f(x,y)$ that is a product of two inhomogeneous linear forms on a lattice in $\mathbb{R}^2$ appears in E. S. Barnes, "The inhomogeneous minima of binary quadratic forms (IV)", *Acta Math.* **92** (1954) 235–264. Although this is the fourth paper in a series, the first three were written with H. P. F. Swinnerton-Dyer, and the use of divided cells only appeared in the third paper. Thus, this paper can be considered to be the beginning of the systematic study.

Subsequent work appeared in the dissertations of Peter Blanksby and Jane Pitman, written under the direction of Barnes. One of the results of Pitman is a striking proof of the existence of divided cells and a relation of this construction to the ordinary continued fraction algorithm. The significance of this construction emerged in conversations with David Crisp, William Moran, and Charles Pearce at the University of Adelaide. Our definitions differ slightly from those of Barnes since we want to allow degenerate cases.

Definition. *A divided cell of a lattice with respect to $f(x,y)$ is a parallelogram with lattice points as vertices that is fundamental domain of the lattice such that each quadrant into which the plane is divided by the lines on which the linear forms vanish contains a vertex of the cell.*

A line on which one of the factors of the expression vanishes will be called an **axis** of the expression. Corresponding to the order of the factors, there is a **first axis** and a **second axis**.

Divided cells of a given expression (once they have been shown to exist) form a **chain** in which a step in one direction is given by extending the sides crossing the first axis, extending them to lines, and marking the lattice intervals on that line which cross the second axis. Since divided cells cannot contain lattice points in their interior, the vertices of the cells must be vertices of the convex hulls of the lattice points in each quadrant (which can be called the **Klein polygon** to recognize its use by Felix Klein to describe the continued fraction in the homogeneous case). If the vertex in a given quadrant changes in a step, it can only be replaced by the other endpoint of an edge of the Klein polygon in the direction in which the second factor decreases.

One approach to inhomogeneous approximation would be to insist that the lattice be the standard integer lattice. In this picture, the parameters of the problem are the six coefficients of the linear factors of $f(x,y)$. The values of these coefficients are naturally arranged in a 3 by 2 matrix. Then $f(x,y)$ is the product of the entries of the product of this matrix with a column whose entries are $(x,y,1)$. A change of basis multiplies this column **on the left** by the 3 by 3 matrix of an **affine** change of coordinates. The upper left 2 by 2 submatrix of this matrix gives a homogeneous change of basis, so it has determinant ± 1. The third row must be $[0,0,1]$ to preserve **columns** of the form $(x,y,1)$. The other elements in the third column are integers representing translation. Problems may be formulated in any coordinates, but should be transformed to make the lattice into an integer lattice. We assume this has been done. Subsequent changes of coordinates will use integer matrices of determinant ± 1. A change of basis in the lattice gives an equivalence relation on the matrices used to describe $f(x,y)$ in which that matrix is multiplied **on the right** by the matrix expressing the change of coordinates.

It is convenient to allow the coefficients in each factor may be scaled. The only effect of scaling one factor is to scale the value of the expression by the same amount, so the **ratio** of the value to a two-by-two determinant of coefficients should be considered. The determinant of the first two columns is unchanged by affine changes in the description of the lattice, so this is the appropriate quantity to which the value of the expression should be compared. It will be convenient to normalize the matrix to have the a special property of the matrix of the first two columns that describes the **shape** of the cell rather than its **size**, so the determinant of this matrix will not take a fixed value. This will be called the **discriminant** of the expression. (There is a similar convention in the homogeneous case where the determinant is compared to the value of the form at $(1, 0)$.)

Each divided-cell step gives an affine change of basis in the lattice.

At this point, the divided-cell algorithm is described in terms of operations on a matrix relating the expression to the lattice, so only the relation of the basis of the lattice to the axes of the expression is relevant, so the coordinates in space can be changed to put the axes in standard position and move the lattice.

2. Reduced matrices

A divided cell is given by identifying the vertex at which both factors are negative and the **displacements** corresponding to the sides of the parallelogram. This point will be taken as the origin and the lattice vertices of the cell will be represented by the matrix

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

where the signs of the value matrix equal to the product of the matrix representing $f(x, y)$ with this matrix is

$$\begin{bmatrix} - & + & - & + \\ - & - & + & + \end{bmatrix}.$$

Left multiplication by a diagonal matrix with positive quantities on the diagonal does not change this pattern.

It is convenient to also allow any of these quantities to be zero. This corresponds to cases in which $f(x, y) = 0$ on the lattice. This is treated as a special case, and usually excluded. However, including this possibility gives a **compact set** of representatives of divided cell configurations. The “liminf” appearing as a measure of the fit between a lattice and the axes of $f(x, y)$ can be realized as a true infimum.

If the matrix of $f(x, y)$ is

$$M_f = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

with respect to a **divided-cell basis** of the lattice, then $c \le 0$, $f \le 0$, $a + c \ge 0$, $d + f \le 0$, $b + c \le 0$, $e + f \ge 0$, $a + b + c \ge 0$, $d + e + f \ge 0$.

Barnes calls a two by two matrix **I-reduced** if a third column can be found satisfying these inequalities. This allows much of the effort to be devoted to the first two columns of M_f. It is an easy consequence of the inequalities already noted that this matrix is I-reduced if and only if $a \ge |b|$ and $e \ge |d|$. It is convenient to scale the rows of M_f so that $a = e = 1$.

The first two columns of M_f represent the quadratic terms in $f(x, y)$. If b and d have opposite signs, this quadratic form is **reduced in the sense of Gauß** and M_f is said to be **G-reduced**.

If the quadratic part of $f(x, y)$ is a multiple of a form with integer coefficients, there are only finitely many equivalent G-reduced forms under changes of lattice basis, and they are easily characterized. Barnes extended (in Lemma 2.1) the one characterization of G-reduced forms to a characterization of I-reduced forms that shows that there are only finitely many I-reduced forms equivalent to a given form with integer coefficients. We repeat his analysis, which excludes forms that represent zero, for the convenience of the reader.

Let $Ax^2 + Bxy + Cy^2$ be an I-reduced form with integer coefficients equivalent to a given form. Then $\Delta^2 = B^2 - 4AC$ is the discriminant of the form, which will be the same for all equivalent forms. If $B > 0$, the factorization of $f(x, y)$ and definition of I-reduced give

$$|B - \Delta| < 2\,|A|\,, \quad |B + \Delta| > 2\,|C| \tag{I}$$

Note that $\Delta^2 - B^2 = -4AC$ so, if $B < \Delta$, we are in the G-reduced case, and this reduces to testing the finitely many factorizations of $\Delta^2 - B^2$ for finitely many B. If $B > \Delta$, A and C have the same sign and it is convenient to introduce $K = C - A$. Then (I) implies

$$|K| = |C - A| < \frac{1}{2}(B + \Delta) - \frac{1}{2}(B - \Delta) = \Delta.$$

Also,

$$(A + C)^2 - K^2 = 4AC = B^2 - \Delta^2,$$

which leads to

$$\Delta^2 - K^2 = (B - A - C)(B + A + C).$$

Again, there are finitely many K to consider and finitely many factorizations to be tested, each of which gives a system uniquely determining A, B, and C.

3. An example

Everyone's favorite example is $\Delta^2 = 5$. Since $4|(\Delta^2 - B^2)$, B must be odd. In the G-reduced case, this allows only $B = 1$ and then $AC = -1$, so $-C = A = \pm 1$.

The remaining case is almost as easy. For all odd B, $\Delta^2 - B^2 \equiv 4 \pmod 8$, so A and C must both be odd, and K even. Only $K = 0$ and $K = \pm 2$ need to be considered, giving $\Delta^2 - K^2$ equal to 5 or 1, respectively. However, the latter case leads to $A = -C$, contradicting the requirement that they have the same sign. Thus, only $K = 0$ is allowed, and this leads to $A = C = 1$ and $B = 3$ (along with those changes of sign that need to be considered).

4. Pitman's theorem

The possible third columns building M_f from an I-reduced matrix lie in a rectangle. In the case of a G-reduced matrix, the remainder of the cell lies in the rectangles of I-reduced matrices that are **neighbors** of the G-reduced matrix. In particular, every G-reduced matrix equivalent to the matrix of the first two columns of M_f — which can be found using the ordinary continued fraction — can be used to find an I-reduced matrix whose allowed third column gives a matrix equivalent to M_f. This gives an **algebraic** proof of the existence of divided cells, and hints at what might be a suitable replacement in higher dimensions.

In the case where $b < 0$ and $d > 0$, the neighbors are

$$\begin{bmatrix} a & a + b \\ d & d + e \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a - b & b \\ d - e & e \end{bmatrix}.$$

In the example of the previous section, there is only one G-reduced form, and the two neighbors are versions of the one other I-reduced form.

Other cases are likely to have additional I-reduced forms.

5. The chain of cells

Suppose now that we have a divided cell. Thus the first two columns of M_f are I-reduced, and the third column satisfies additional inequalities. If $b = 0$, we have reached a terminal state where the cell has a side parallel to the second axis of $f(x, y)$. Otherwise, the next cell in the chain has its first column equal to a multiple of the second column of this matrix, and this multiple must have the same sign as b. If $b > 0$, the **top** of the cell is on the line through (c, f), and the **bottom** of the cell is on the line through $(a + c, d + f)$; if $b < 0$, the **bottom** of the cell is on the line through (c, f), and the **top** of the cell is on the line through $(a + c, d + f)$.

Consider the case in which $b > 0$. If there are h forward steps along the top from (c, f) to the new **second quadrant** point, then $h \geq 1$. If there are k backward steps along the bottom from $(a + c, d + f)$ to the new **fourth quadrant** point, then $k \geq 0$. Since the former point is to the left of the second axis and the latter point is to the right, it follows that $(h + k)b \leq a$. Similarly, $(h + k + 2)b \geq a$. Since we know all vertices of the cell, we can see that the new second column is $h + k - 1$ times the original second column minus the original first column. It is easily seen that this is always I-reduced. The vertex in the third quadrant that we are using as our reference point is the sum of the first and third columns minus $k + 1$ times the second column.

Each choice of signs of b and d allows a similar analysis. Early work on this subject used additional symmetries to limit the number of cases, but it is probably better to distinguish four cases while using the symmetry to recognize similarities in the formulas.

For cells that are neighbors of G-reduced cells, a description of the step in terms of the neighboring G-reduced cell should be attempted, since this would limit the number of distinct I-reduced cells that need to be considered.