アルゴリズム 84－4
（2002．5．23）

Matrix Rounding under the L_p-Discrepancy Measure and Its Application to Digital Halftoning

浅野哲夫[1], 加藤直樹[2], 小保方幸次[1], 徳山豪[3]

[1] 北陸先端科学技術大学院大学
[2] 京都大学大学院工学系研究科
[3] 東北大学大学院情報科学研究科

本論文では，L_p-ディスクレパンシーが最小になるように実数値行列を２値行列に丸める問題を扱う．L_p-ディスクレパンシーの基準を定義するために，行列上の領域（部分行列）の族 $\mathcal{F}$ を導入し，これを用いてハイパーグラフを考える．本論文では，上記の最適化問題の計算複雑度が領域族 $\mathcal{F}$ の選び方によってどのように変わるかを考察する．

Matrix Rounding under the L_p-Discrepancy Measure and Its Application to Digital Halftoning

Tetsuo Asano[1], Naoki Katoh[2], Koji Obokata[1], and Takeshi Tokuyama[3]

[1]School of Information Science, Japan Advanced Institute of Science and Technology, Tatsunokuchi, Japan.
[2]Graduate School of Engineering, Kyoto University, Kyoto, Japan.
[3]Graduate School of Information Sciences, Tohoku University, Sendai, Japan.

In this paper we study the problem of rounding a real-valued matrix into an integer-valued matrix to minimize an L_p-discrepancy measure between them. To define the L_p-discrepancy measure, we introduce a family $\mathcal{F}$ of regions (rigid submatrices) of the matrix, and consider a hypergraph defined by the family. This paper shows how the difficulty of the problem depends on the choice of the region family $\mathcal{F}$.

1 Introduction

Rounding is an important operation in numerical computation, and plays key roles in digitization of analogue data. Rounding of a real number a is basically a simple problem: We round it to either $\lfloor a \rfloor$ or $\lceil a \rceil$, and we usually choose the one nearer to a. However, we often encounter a data consisting of more than one real numbers instead of a singleton. If it has n numbers, we have 2^n choices for rounding since each number is rounded into either its floor or ceiling. If the original data set has some feature, we need to choose a rounding so that the rounded result inherits as much of the feature as possible. The feature is described by using some combinatorial structure; we indeed consider a hypergraph $\mathcal{H}$ on the set. A typical input set is a multi-dimensional array of real numbers, and we consider a hypergraph whose hyperedges are its subarrays with contiguous indices. In this paper, we focus on two-dimensional arrays; In other words, we consider rounding problems on matrices.

1.1 Rounding and discrepancy

Given an $N \times N$ matrix $A = (a_{ij})_{1 \le i,j \le N}$ of real numbers, its rounding is a matrix $B = (b_{ij})_{1 \le i,j \le N}$ of integral values such that b_{ij} is either $\lfloor a_{ij} \rfloor$ or $\lceil a_{ij} \rceil$ for each (i,j). There are 2^{N^2} possible roundings of a given A, and we would like to find an optimal rounding with respect to a given criterion. This is called the *matrix rounding problem.* Without loss of generality, we can assume that each entry of A is in the closed interval $[0,1]$ and each entry is rounded to either 0 or 1.

In order to give a criterion to determine the quality of roundings, we define a distance in the space of all $[0,1]$-valued $N \times N$ matrices. We introduce a family $\mathcal{F}$ of regions over the $N \times N$ integer grid

$$\begin{aligned} G_n &= \{(1,1),(1,2)\ldots,(N,N)\} \\ &= \{p_1,p_2,\ldots,p_n\}, \quad n=N^2. \end{aligned}$$

This means that each entry location of a matrix is denoted by a symbol p_i and that the order $(p_1,\ldots,p_n)$ is arbitrary. Let $\mathcal{A} = \mathcal{A}(G_n)$ be the space of all $[0,1]$-valued matrices with the index set G_n, and let $\mathcal{B} = \mathcal{B}(G_n)$ be its subset consisting of all $\{0,1\}$-valued matrices. Let R be a region in $\mathcal{F}$. † For an element $A \in \mathcal{A}$, let $A(R)$ be the sum of entries of A located in the region R, that is,

$$A(R) = \sum_{p_i \in R} a_{p_i}.$$

We define a distance $Dist_p^{\mathcal{F}}(A,A')$ between two elements A and A' in $\mathcal{A}$ for a positive integer p by

$$Dist_p^{\mathcal{F}}(A,A') = [\sum_{R\in\mathcal{F}} |A(R)-A'(R)|^p]^{1/p}.$$

The distance is called the L_p-distance with respect to $\mathcal{F}$. The L_∞ distance with respect to $\mathcal{F}$ is defined by

$Dist_\infty^{\mathcal{F}}(A,A') = \lim_{p\to\infty} Dist_p^{\mathcal{F}}(A,A')$
$= \max_{R\in\mathcal{F}} |A(R)-A'(R)|$.

Using the notations above, we can formally define the matrix rounding problem:

L_p-Optimal Matrix Rounding Problem: $\mathcal{P}(G_n,\mathcal{F},p)$: Given a $[0,1]$-matrix $A \in \mathcal{A}$, a family $\mathcal{F}$ of subsets of G_n, and a positive integer p, find a $\{0,1\}$-matrix $B \in \mathcal{B}$ that minimizes

$$Dist_p^{\mathcal{F}}(A,B) = [\sum_{R\in\mathcal{F}} |A(R)-B(R)|^p]^{1/p}.$$

Also, we are interested in the following combinatorial problem:

L_p-Discrepancy Bound:
Given a $[0,1]$-matrix $A \in \mathcal{A}$, a family $\mathcal{F}$ of subsets of G_n, and a positive integer p, investigate upper and lower bounds of

$$\mathcal{D}(G_n,\mathcal{F},p) = \sup_{A\in\mathcal{A}} \min_{B\in\mathcal{B}} Dist_p^{\mathcal{F}}(A,B).$$

The pair $(G_n,\mathcal{F})$ defines a hypergraph on G_n, and $\mathcal{D}(G_n,\mathcal{F},\infty)$ is called the *inhomogeneous discrepancy* of the hypergraph [6]. Abusing the notation, we call $\mathcal{D}(G_n,\mathcal{F},p)$ the (inhomogeneous) L_p-discrepancy of the hypergraph, and also often call $Dist_p^{\mathcal{F}}(A,B)$ the L_p-discrepancy measure of (quality of) the output B with respect to $\mathcal{F}$.

1.2 Motivation and our application

The most popular example of the family $\mathcal{F}$ is the set of all rectangular subregions in G_n (i.e. the set of all rigid submatrices), and the corresponding L_∞-discrepancy measure is utilized in many application areas such as Monte Carlo simulation and computational geometry. However, unfortunately, if we consider the family of all rectangular subregions, the discrepancy bound (for the L_∞ measure) is known to be large ($\Omega(\log n)$ and $O(\log^3 n)$). See Beck and Sós's survey [6] for the theory. Moreover, the optimal matrix rounding problem becomes NP-hard [2].

Therefore, we would like to seek for a family for which a rounding with low discrepancy can be computed in polynomial time as well as the low-discrepancy rounding is useful in an important application. Also, L_∞-discrepancy measure is not suitable for some applications, and we often need L_p-discrepancy measure (typically for $p=1$ or 2). For the purpose, we investigate (1) how a geometric structure of a family of regions reflects the combinatorial

† Strictly speaking, R can be any subset of G_n. Although we implicitly assume that R forms some connected portion on the grid G_n, the connectivity assumption is not used throughout the paper.

discrepancy bound and the computational difficulty of the matrix rounding problem, and (2) how we can establish a theory for the L_p measure.

In particular, we focus on the digital halftoning application of the matrix rounding problem, where we should consider smaller families of rectangular subregions as $\mathcal{F}$. More precisely, the input matrix represents a digital (gray) image, where a_{ij} represents the brightness level of the (i,j)-pixel in the $N \times N$ pixel grid. Typically, N is between 256 and 4096, and a_{ij} is an integral multiple of $1/256$: This means that we use 256 brightness levels. If we want to send an image using fax or print it our by a dot (or ink-jet) printer, brightness levels available are limited. Instead, we replace A by an integral matrix B so that each pixel uses only two brightness levels. Here, it is important that B looks similar to A; in other words, B should be a good approximation of A.

For each pixel $p = (i,j)$, if the average brightness level of B in each of its neighborhoods (regions containing p in a suitable family of regions) is similar to that of A, we can expect that B is a good approximation of A. For this purpose, the set of all rectangles is not suitable (i.e., it is too large), and we may use a more compact family. Moreover, since human vision detects global features, the L_1 or L_2 measure should be better than the L_∞ measure to obtain a clear output image. This intuition is supported by our experimental results; for example, edges of objects are often blurred in the output based on the L_∞-discrepancy measure, while they are sharply displayed if we use the L_1-discrepancy measure.

1.3 Known results on L_∞ measure

For the L_∞ measure, the following beautiful combinatorial result is classically known:

Theorem 1.1 [Baranyai[5] 1974] *Given a real-valued matrix $A = (a_{ij})$ and a family $\mathcal{F}$ of regions consisting of all rows, all columns and the whole matrix, there exists an integer-valued matrix $B = (b_{ij})$ such that*
$|\sum_{(i,j)\in R} a_{ij} - \sum_{(i,j)\in R} b_{ij}| < 1$
holds for every $R \in \mathcal{F}$.

Translating the theorem in our terminologies, the L_∞-discrepancy of the matrix rounding problem for the family of regions consisting of all rows, all columns, and the whole matrix is bounded by 1. Also, the combinatorial structure and algorithmic aspects of roundings of (one-dimensional) sequences with respect to the L_∞-discrepancy measure are investigated in recent studies [2, 18].

The *constraint matrix* $\mathcal{C}(G_n, \mathcal{F}) = (c_{ij})_{p_i \in G_n, R_j \in \mathcal{F}}$ of the hypergraph $(G_n, \mathcal{F})$ is defined by $c_{ij} = 1$ if $p_i \in R_j$ and 0 otherwise. A hypergraph is called *unimodular* if its constraint matrix is totally unimodular, where a matrix C is *totally unimodular* if the determinant of each square submatrix of C is equal to $0, 1$, or -1.

Both the Baranyai's problem and the sequence rounding problems correspond to rounding problems with respect to totally unimodular hypergraphs. The L_∞-discrepancy problem can be formulated as an integer programming problem, and the unimodularity implies that its relaxation has an integral solution. A classical theorem of Ghouila-Houri [10] implies that totally unimodularity is a necessary and sufficient condition for the existence of a rounding with a L_∞ discrepancy less than 1. Moreover, the following sharpened result is given by Doerr [7]:

Theorem 1.2 *If $(G_n, \mathcal{F})$ is a unimodular hypergraph, there exists a rounding $B = (b_{ij})$ of $A = (a_{ij})$ satisfying*

$$|\sum_{(i,j)\in R} a_{ij} - \sum_{(i,j)\in R} b_{ij}| < \min\{1 - \frac{1}{n+1}, 1 - \frac{1}{m}\}$$

for every $R \in \mathcal{F}$, where $m = |\mathcal{F}|$.

This bound is sharp. Moreover, L_∞-optimal rounding can be computed in polynomial time if $\mathcal{F}$ is totally unimodular.

1.4 Our results

We would like to consider L_p-discrepancy measure instead of L_∞-discrepancy measure. If the hypergraph is totally unimodular, an $|\mathcal{F}|^{1/p}$ upper bound for the L_p-discrepancy can be derived from Theorem 1.2 trivially. We first improve the upper bound to $\frac{1}{2}|\mathcal{F}|^{1/p}$ for $p \leq 3$, and show that the bound is tight. We also consider the family $\mathcal{F}$ consisting of all 2-by-2 rigid submatrices, for which the matrix rounding problem is known to be NP-hard [2] (accordingly, the family is not totally unimodular). The L_∞-discrepancy of this family is known to be at most $\frac{5}{3}$ [3] and at least 1; Hence, $\frac{5}{3}|\mathcal{F}|$ is a trivial upper bound for the L_1-discrepancy. We improve the upper bound to $\frac{3}{4}|\mathcal{F}|$, and also design an approximation algorithm attaining $\frac{9}{16}|\mathcal{F}|$ provable error bound measured from the optimal solution.

Next, we consider the optimization problem. If the hypergraph is totally unimodular, the rounding minimizing the L_p-discrepancy can be computed in polynomial time by translating it to a separable convex programming problem and applying known general algorithms [11, 12]. However, we want to define a class of region families for which we can compute the optimal solution more efficiently, as well as the class is useful in applications (in particular, the digital halftoning application). We consider a union of two laminar families (defined in Section 3), and show that the matrix rounding problem can be formulated into a minimum cost flow problem, and hence solved in polynomial time. Finally, we implemented the algorithm using LEDA[14]. Some output pictures of the algorithm applying to the digital halftoning problem are included.

2 Mathematical programming

2.1 Linear convex program

We give a formulation of the L_p-discrepancy problem into an integer convex programming problem where the objective function is a separable convex function, i.e., a sum of univariate convex functions.

Introducing a new variable $y_i = B(R_i) = \sum_{p_j \in R_i} b_{p_j}$ for each $R_i \in \mathcal{F}$, the problem $\mathcal{P}(G_n, \mathcal{F}, p)$ is described in the following form:

(P1): minimize $[\sum_{R_i \in \mathcal{F}} |y_i - A(R_i)|^p]^{1/p}$
subject to $y_i = \sum_{p_j \in R_i} b_{p_j}$,
$i = 1, \ldots, m = |\mathcal{F}|$
and $b_{p_j} = 0, 1,\ j = 1, \ldots, n = |G_n|$.

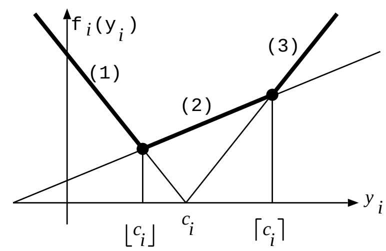

Figure 1: Conversion of the convex objective function into a piecewise linear convex function.

When $p < \infty$, the objective function can be replaced with $\sum_{R_i \in \mathcal{F}} |y_i - c_i|^p$, where $c_i = A(R_i) = \sum_{p_j \in R_i} a_{p_j}$ is a constant depending only on input values. Now, $|y_i - c_i|^p$ is a convex function independent of other y_js. The constraints $y_i = \sum_{p_j \in R_i} b_{p_j},\ i = 1, \ldots, m$ are represented by $(-I, \mathcal{C}(G_n, \mathcal{F}))Y = 0$ using the constraint matrix where
$Y = (y_1, \ldots, y_m, b_{p_1}, \ldots, b_{p_n})^T$
and I is an identity matrix.

Although the objective function is now a separable convex function, its nonlinearity gives difficulty to analyze the properties of the solution. Thus, we apply the idea of Hochbaum and Shanthikumar [11] to replace $|y_i - c_i|^p$ with a piecewise linear convex continuous function $f_i(y_i)$ which is equal to $|y_i - c_i|^p$ for each integral value of y_i in $[0, |R_i|]$. This is because we only need integral solutions, and if each b_{p_j} is integral, y_i must be a nonnegative integer less than or equal to $|R_i|$. Typically for $p = 1$, $f_i(y_i)$ is illustrated in Figure 1.

Thus, we obtain the following problem (P2):

(P2): minimize $\sum_{R_i \in \mathcal{F}} f_i(y_i)$
subject to $y_i = \sum_{p_j \in R_i} b_{p_j}$,
$i = 1, \ldots, m = |\mathcal{F}|$
and $b_{p_j} = 0, 1,\ j = 1, \ldots, n = |G_n|$.

Thus, we can formulate the problem into a integer programming problem where the objective function is a separable piecewise-linear convex function.

2.2 Total unimodularity

Let (P3) be the continuous relaxation obtained from (P2) by replacing the integral condition of b_{p_j} with the condition $0 \leq b_{p_j} \leq 1$. Note that this is different from the continuous relaxation of (P1), since the objective function of (P2) is larger than that of (P1) at non-integral values.

If the matrix is totally unimodular, it is well-known that (P3) has an integral solution. This is a key to derive discrepancy bounds and also algorithms.

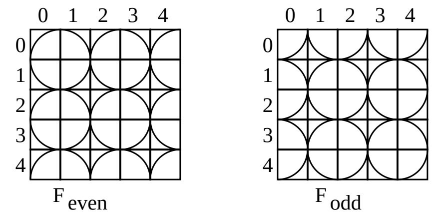

Figure 2: 2-partition family of 2×2 regions.

3 Geometric region families

In this section we consider interesting classes of families whose associated constraint matrices are totally unimodular. We call such a family a *unimodular family*, since associated hypergraph is unimodular.

A family $\mathcal{F} = \{R_1, R_2, \ldots, R_m\}$ is a *partition family* (or a partition) of G_n if $\bigcup_{i=1}^m R_i = G_n$ and $R_i \cap R_j = \emptyset$ for any $R_i \neq R_j$ in $\mathcal{F}$. A *k-partition family* is a family of regions on a matrix which is a union of k different partitions of G_n.

A family $\mathcal{F}$ of regions on a grid G_n is a *laminar family* if one of the following holds for any pair R_i and R_j in $\mathcal{F}$: (1) $R_i \cap R_j = \emptyset$, (2) $R_i \subset R_j$ and (3) $R_j \subset R_i$. The family is also called a laminar decomposition of the grid G_n. In general, a *k-laminar family* is a family of regions on a matrix which is a union of k different laminar families.

Proposition 3.1 *A 2-laminar family is unimodular.*

Direct applications of Proposition 3.1 lead to various unimodular families of regions. The family of regions defined in Baranyai's theorem is a 2-laminar family. Also, take any 2−partition family consisting of 2×2 regions on a matrix. For example, take all 2×2 regions with their upper left corners located in even points (where the sums of their row and column indices are even). The set of all those regions defines two partition families $\mathcal{F}_{even}$ and $\mathcal{F}_{odd}$ where $\mathcal{F}_{even}$ (resp. $\mathcal{F}_{odd}$) consists of all 2×2 squares with their upper left corners lying at even (resp. odd) rows (see Figure 2). Since any 2×2 region is partitioned into two 2×1 regions, with these 2×2 regions and 2×1 regions we have a 2-laminar family. This kind of families plays an important role in Section 4.2 and also in our experiment.

A 3−partition family is not unimodular in general. However, there are some families which are not 2-laminar but unimodular: For example, the set of all rectangular rigid submatrices of size 2 (i.e., domino tiles) is a 4-partition family, but it is unimodular.

4 Upper bounds

4.1 Bound for unimodular hypergraphs

In this subsection, we prove the following theorem for the L_p-discrepancy of a unimodular family.

Theorem 4.1 *If $\mathcal{F}$ is unimodular and $p \leq 3$, for any $A \in \mathcal{A}$ we have*

$$\min_{B \in \mathcal{B}} Dist_p^{\mathcal{F}}(A, B) \leq \frac{1}{2}|\mathcal{F}|^{1/p}.$$

It is easy to give an instance to show that the bound is tight: Consider Baranyai's problem on a matrix having $\frac{1}{2}$ entries in its diagonal position (other entries are zeros).

For the case $p > 3$, we have the following:

Theorem 4.2 *If $\mathcal{F}$ is unimodular and $p > 3$, for any $A \in \mathcal{A}$ we have*

$$\begin{aligned} & \min_{B \in \mathcal{B}} Dist_p^{\mathcal{F}}(A, B) \\ \leq \quad & \left(\frac{p^p}{(p+1)^{p+1}} + \frac{2^p(p-1)}{(p+1)^{p+1}}\right)^{1/p} |\mathcal{F}|^{1/p}. \end{aligned}$$

The term $(p^p/(p+1)^{p+1} + 2^p(p-1)/(p+1)^{p+1})^{1/p}$ is 0.550 and 0.587 if $p = 4$ and $p = 5$, respectively, and it is always less than $p/(p+1)$.

4.2 Some special case

The method in the previous subsection does not work for a non-unimodular case. A simple but interesting family defining a non-unimodular hypergraph is the family of all 2×2 regions of A. The known upper bound is merely $\frac{5}{3}|\mathcal{F}|^{1/p}$ [3]. We obtain the following result:

Theorem 4.3 *For any $A \in \mathcal{A}(G_n)$, and a family $\mathcal{F}$ of 2×2 regions of the matrix, (1) we have*

$$\min_{B \in \mathcal{B}} Dist_1^{\mathcal{F}}(A, B) \leq \frac{3}{4}|\mathcal{F}|,$$

and (2) we can find in polynomial time a $\{0,1\}$-valued matrix B such that
$\sum_{R\in\mathcal{F}} |A(R) - B(R)|$
$\leq \sum_{R\in\mathcal{F}} |A(R) - B^*(R)| + \frac{9}{16}|\mathcal{F}|$,
where B^* is the (unknown) optimal solution.

5 Optimal rounding

The following theorem [11] is known for a totally unimodular convex programming:

Theorem 5.1 *Nonlinear separable convex optimization problem $\min\{\sum_{i=1}^{n} f_i(x_i) \mid Ax \geq b\}$ on linear constraints with a totally unimodular matrix A can be solved in polynomial time.*

Indeed, we have borrowed techniques from the proof of the above theorem in the previous section. The following is thus immediately obtained:

Corollary 5.2 *Matrix rounding problem $\mathcal{P}(G_n, \mathcal{F}, p)$ for $p < \infty$ is solved in polynomial time in $n = |G_n|$ if its associated constraint matrix $\mathcal{C}(G_n, \mathcal{F})$ is totally unimodular.*

However, we needed a more practical algorithm for our experiments that runs fast for large-scale problem instances. In this section we will show how to solve the matrix rounding problem for a 2-laminar family based on the minimum-cost flow algorithm. For improving the readability, we mainly discuss the case for the L_1-discrepancy measure.

Let $\mathcal{F}$ be 2-laminar family given as a union of laminar families $\mathcal{F}_1$ and $\mathcal{F}_2$. Without loss of generality, we can assume that each pixel is covered by at least one region in each of $\mathcal{F}_1$ and $\mathcal{F}_2$. We define a network $\mathcal{N} = (V, E \cup \tilde{E}')$. Figure 3 illustrates the shape of the network. Nodes of $\mathcal{N}$ are identified to regions of $\mathcal{F}$. Edges between nodes in $\mathcal{F}_1$ form a tree T_1 representing the lattice structure of the laminar family $\mathcal{F}_1$. The tree T_2 for $\mathcal{F}_2$ is defined similarly. The edges are directed from leaves to the root in T_1 while from the root to leaves in T_2. The edge outgoing from a node for R_i in the tree T_1 is denoted by $e(R_i)$, and a unique edge incoming into a node for R_j in the tree T_2 is denoted by $e(R_j)$. The edge set E contains an edge $e(r_1, r_2)$ from the root of T_1 towards T_2 in addition to the above-defined edges. The edges in $\tilde{E}'$ correspond to matrix entries, and each edge is denoted like $e(p_i)$ for $p_i \in G_n$. From the laminar condition, there is a unique minimal region containing p_i in each of $\mathcal{F}_1$ and $\mathcal{F}_2$. The edge $e(p_i)$ connects the corresponding leaf nodes of T_1 and T_2. See Figure 3 for illustration.

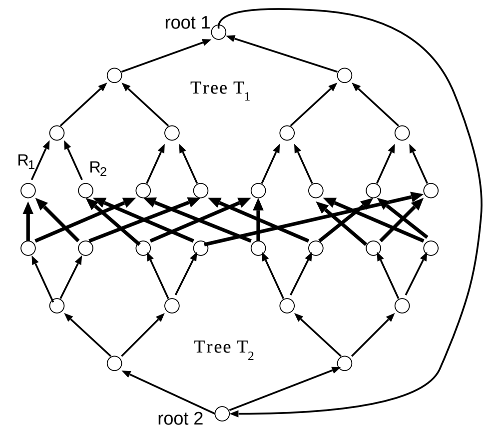

Figure 3: Network $\mathcal{N}$

Now, we define capacity and cost coefficient of each edge. (Lower bound on the flow of each edge is defined to be 0.) Capacity of an edge $e(p_i)$ in $\tilde{E}'$ is determined simply as 1 because the value associated with p_i is restricted to $[0,1]$. Capacity of an edge $e(R_j)$ is given by $|R_j|$ to reflect the size of R_j. Capacity of the edge $e(r_1, r_2)$ is determined as $|G_n|$.

Determining the cost coefficient of an edge $e(R_i)$ is not straightforward although the cost coefficients of $e(p_i)$s and $e(r_1, r_2)$ are defined to be 0s. This is because each term of the objective function depends on the difference between $B(R_j)$ and $A(R_j)$, that is, $|B(R_j) - A(R_j)|$.

Recall the argument in Section 2: To prove the polynomial-time solvability we have introduced a new variable $y_j = B(R_j) = \sum_{p_i \in R_j} b_{p_i}$. $|B(R_j) - A(R_j)|$ is converted into $|y_i - c_i|$, where $c_i = A(R_i) = \sum_{p_j \in R_i} a_{p_j}$ is a constant determined by input values. $|y_i - c_i|$ is further replaced by the piecewise linear convex function $f_i(y_i)$ which coincides with $|y_i - c_i|$ at each integral value of y_i in $[0, |R_i|]$.

To reflect the new form of the objective function, we divide each edge $e(R_i)$ into three parallel edges with different capacity and costs: $e_1(R_i)$ has capacity $\lfloor c_i \rfloor$ and cost $c^1 = -1$. $e_2(R_i)$ has capacity $\lceil c_i \rceil - \lfloor c_i \rfloor$ and cost $c^2 = \lfloor c_i \rfloor + \lceil c_i \rceil - 2c_i$. For the third edge $e_3(R_i)$, its capacity is ∞ and its cost is $c^3 = 1$. Since $c^1 \le c^2 \le c^3$, to minimize the overall cost for these three edges the flow at $e_2(R_i)$ is zero unless the first edge $e_1(R_i)$ is full, that is, the flow at $e_1(R_i)$ is $\lfloor c_i \rfloor$. Similarly, flow at $e_3(R_i)$ is positive only if the two edges $e_1(R_i)$ and $e_2(R_i)$ are both full.

The cost associated with an edge is determined by multiplying the above coefficient to the flow in the edge. When the total amount of flow in the three edges is given by y_i, the total cost is given by $f_i(y_i) - c_i$ in any case (see, e.g., Ahuja-Magananti-Orlin[1]). Since c_i is a constant, the constant term does not affect the optimality.

Theorem 5.3 *Given a* $[0,1]$*-matrix* A *and a* 2*-laminar family* $\mathcal{F}$*, an optimal binary matrix* B *that minimizes the distance* $Dist_1^{\mathcal{F}}(A, B)$ *is computed in* $O(n^2 \log^2 n)$ *time, where* n *is the number of matrix elements.*

Theorem 5.4 *Given a* $[0,1]$*-matrix* A *and a* 2*-laminar family* $\mathcal{F}$*, an optimal binary matrix* B *that minimizes the distance* $Dist_p^{\mathcal{F}}(A, B)$ *(*$p \ge 2$*) is computed in* $O(n^2 \log^3 n)$ *time, where* n *is the number of matrix elements.*

6 Digital Halftoning

The quality of color printers has been drastically improved in recent years, mainly based on the development of fine control mechanism. On the other hand, there seems to be no great invention on the software side of the printing technology. What is required is a technique to convert a continuous-tone image into a binary image consisting of black and white dots so that the binary image looks very similar to the input image. From a theoretical standpoint, the problem is how to approximate an input $[0,1]$-array by a binary array. Since this is one of the central techniques in computer vision and computer graphics, a great number of algorithms have been proposed (see, e.g., [13, 9, 4, 15, 16]). However, there have been very few studies toward the goal of achieving an optimal binary image under some reasonable criterion; maybe because the problem itself is very practically oriented. A popular distortion criterion is the *Frequency Weighted Mean Square Error* (FWMSE) based on the human visual system which is to minimize the sum of the differences of the weighted averages between input and output images. Our discrepancy measure which has been discussed in this paper is a hopeful replacement; Indeed, the L_2-discrepancy measure can be regarded as a simplified version of the FWMSE criterion.

7 Concluding remarks

We have considered the matrix rounding problem based of L_p-discrepancy measure. Although we have shown that the measure is useful in application to the digital halftoning application, the current algorithm is too slow if we want to require speed together with the high-quality requirement. The problem comes from the quadratic time complexity. It is desired to design a faster algorithm (even an approximation algorithm). Moreover, it is an interesting question to investigate what kind of region families give the best criterion for the halftoning application. Once we know such a region family, it is valuable to design an algorithm (heuristic algorithm if the problem for solving the optimal solution is intractable) for the criterion.

Acknowledgment

The authors would like to thank Tomomi

Matsui, Koji Nakano, and Hiroshi Nagamochi for their valuable comments and helpful discussions.

References

[1] R. Ahuja, T. Magnanti, and J. Orlin: *Network Flows, Theory Algorithms and Applications,* Princeton Hall, 1993.

[2] T. Asano, T. Matsui, and T. Tokuyama: "Optimal Roundings of Sequences and Matrices," *Nordic Journal of Computing*, Vol.7, No.3, pp.241-256, Fall 2000.

[3] T. Asano and T. Tokuyama: "How to Color a Checkerboard with a Given Distribution — Matrix Rounding Achieving Low 2×2- Discrepancy," *to appear in Proc. ISAAC01* (2001).

[4] B. E. Bayer: "An optimum method for two-level rendition of continuous-tone pictures," *Conference Record, IEEE International Conf. on Communications,* 1 (1973), pp.(26-11)–(26-15).

[5] Z. Baranyai: "On the factorization of the complete uniform hypergraphs", in *Infinite and Finite Sets*, eds. A. Hanaj, R. Rado and V. T. Sós, *Colloq. Math. Soc. J'anos Bolyai* **10**, pp.91-108.

[6] J. Beck and V. T. Sós: *Discrepancy Theory*, in *Handbook of Combinatorics* Volume II, (ed. R.Graham, M. Grötschel, and L Lovász) 1995, Elsevier.

[7] B. Doerr: "Lattice Approximation and Linear Discrepancy of Totally Unimodular Matrices–Extended Abstract–," *SIAM-ACM Symposium on Discrete Algorithms*, pp.119-125, 2001.

[8] J. Edmonds and R.M. Karp: "Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems," *Journal of ACM*, 19, 248-264, 1972.

[9] R. W. Floyd and L. Steinberg: "An adaptive algorithm for spatial gray scale," *SID 75 Digest, Society for Information Display*, (1975) pp.36–37.

[10] A. Ghoulia-Houri, "Characterisation des matrices totalement unimodulaires," *C.R. Acad/ Sci. Paris,* 254, (1962) 1192-1194.

[11] D.S. Hochbaum and J.G. Shanthikumar: "Nonlinear separable optimization is not much harder than linear optimization," *Journal of ACM*, Vol.37, No. 4 (1990) 843-862.

[12] A.V. Karzanov and S.T. McCormick: "Polynomial methods for separable convex optimization in totally unimodular linear spaces with applications," *SIAM J. on Computing*, Vol.26, No.4 (1997) 1245-1275.

[13] D.E. Knuth: "Digital halftones by dot diffusion," *ACM Trans. Graphics*, 6-4 (1987), pp.245–273.

[14] The LEDA homepage: http://www.algorithmic-solutions.com/as_html/products/products.html.

[15] J. O. Limb: "Design of dither waveforms for quantized visual signals," *Bell Syst. Tech. J.,* 48-7 (1969), pp. 2555–2582.

[16] B. Lippel and M. Kurland: "The effect of dither on luminance quantization of pictures," *IEEE Trans. Commun. Tech.,* COM-19 (1971), pp.879-888.

[17] M. Minoux: "Solving Integer Minimum Cost Flows with Separable Cost Objective Polynomially," *Mathematical Programming Study* 26 (1986) pp.237-239.

[18] K. Sadakane, N. Takki-Chebihi, and T. Tokuyama: "Combinatorics and Algorithms on Low-Discrepancy Roundings of a Real Sequence", *Proc. 28th ICALP, Springer LNCS2076*, (2001) pp. 166-177.

[19] A.Schrijver: "Theory of Linear and Integer Programming," *Wiley-Interscience Series in Discrete Mathematics*, John Wiley and Sons (1986).