Approximation Algorithms for the Metric Labeling Problem via a New Linear Programming Formulation

Chandra Chekuri* Sanjeev Khanna † Joseph (Seffi) Naor‡ Leonid Zosin§

September 22, 2000

Abstract

We consider approximation algorithms for the metric labeling problem. Informally speaking, we are given a weighted graph that specifies relations between pairs of objects drawn from a given set of objects. The goal is to find a minimum cost labeling of these objects where the cost of a labeling is determined by the pairwise relations between the objects and a distance function on labels; the distance function is assumed to be a metric. Each object also incurs an assignment cost that is label, and vertex dependent. The problem was introduced in a recent paper by Kleinberg and Tardos [19], and captures many classification problems that arise in computer vision and related fields. They gave an $O(\log k \log\log k)$ approximation for the general case where k is the number of labels and a 2-approximation for the uniform metric case. More recently, Gupta and Tardos [14] gave a 4-approximation for the truncated linear metric, a natural non-uniform metric motivated by practical applications to image restoration and visual correspondence.

In this paper we introduce a new natural integer programming formulation and show that the integrality gap of its linear relaxation either matches or improves the ratios known for several cases of the metric labeling problem studied until now, providing a unified approach to solving them. Specifically, we show that the integrality gap of our LP is bounded by $O(\log k \log\log k)$ for general metric and 2 for the uniform metric thus matching the ratios in [19]. We also develop an algorithm based on our LP that achieves a ratio of $2 + \sqrt{2} \simeq 3.414$ for the truncated linear metric improving the ratio provided by [14]. Our algorithm uses the fact that the integrality gap of our LP is 1 on a linear metric. We believe that our formulation has the potential to provide improved approximation algorithms for the general case and other useful special cases. Finally, our formulation admits general non-metric distance functions. This leads to a non-trivial approximation guarantee for a non-metric case that arises in practice [20], namely the truncated quadratic distance function. We note here that there are non-metric distance functions for which no bounded approximation ratio is achievable.

*Bell Labs, 600 Mountain Ave, Murray Hill, NJ 07974. E-mail: chekuri@research.bell-labs.com.

†Dept. of Computer & Information Science, University of Pennsylvania, Philadelphia, PA 19104. E-mail: sanjeev@cis.upenn.edu. Supported in part by an Alfred P. Sloan Research Fellowship.

‡Bell Labs, 600 Mountain Ave, Murray Hill, NJ 07974. On leave from the Computer Science Dept., Technion, Haifa 32000, Israel. E-mail: naor@research.bell-labs.com.

§ITG Inc. This work done while at NEC Research Institute, Inc., 4 Independence Way, Princeton, NJ 08540.

1 Introduction

Motivated by certain classification problems that arise in computer vision and related fields, Kleinberg and Tardos [19] recently introduced the metric labeling problem. In a typical classification problem, one wishes to assign labels to a set of objects to optimize some measure of the quality of the labeling. The metric labeling problem captures a broad range of classification problems where the quality of a labeling depends on the pairwise relations between the underlying set of objects. More precisely, the task is to classify a set V of n objects by assigning to each object a label from a set L of labels. The pairwise relationships between the objects are represented by a weighted graph $G = (V, E)$ where $w(u, v)$ represents the strength of the relationship between u and v. The objective is to find a labeling, a function $f : V \to L$ that maps objects to labels, where the cost of f, denoted by $Q(f)$, has two components.

- For each $u \in V$, there is a non-negative assignment cost $c(u, i)$ to label u with i. This cost reflects the relative likelihood of assigning labels to u.
- For each pair of objects u and v, the edge weight $w(u, v)$ measures the strength of their relationship. The assumption is that strongly related objects should be assigned labels that are *close*. This is modeled in the objective function by the term $w(u, v) \cdot d(f(u), f(v))$ where $d(\cdot, \cdot)$ is a metric on the labels L.

Thus

$$Q(f) = \sum_{u \in V} c(u, f(u)) + \sum_{(u,v) \in E} w(u, v) \cdot d(f(u), f(v))$$

and the goal is to find a labeling of minimum cost. We remark that if the distance function d is not a metric, then determining whether a graph can be colored by k colors is a special case of the labeling problem.

A prototypical application of the metric labeling problem is the *image restoration* problem in computer vision [3, 4, 5]. In the image restoration problem, the goal is to take an image corrupted by noise and restore it to its "true" version. The image consists of pixels and these are the objects in the classification problem. Each pixel has a discretized intensity value associated with it that is possibly corrupted and we would like to restore it to its true value. Thus the labels correspond to intensity values and the goal is to assign a new intensity to each pixel. Further, neighbouring pixels are assumed to be close to each other, since intensities change smoothly. Thus neighbouring pixels have an edge between with a positive weight (assume a uniform value for concreteness). The original or observed intensities are assumed to be close to true values[1]. The assignment cost is some positive cost associated with changing the intensity from its original value to a new value, the larger the change, the larger the cost. To see how the cost function models the restoration consider a pixel corrupted by noise resulting in it having an observed intensity very different from all its neighbouring pixels. By changing the pixel's intensity we incur a cost of relabeling but that can be offset by the edge cost saved by being closer to its neighbours. The assignment costs weigh the labeling in favor of the original values since most of the intensities are likely to be correct (see also the footnote).

The metric labeling problem naturally arises in many other applications in image processing and computer vision. Researchers in these fields have developed a variety of good heuristics that use classical combinatorial optimation techniques such as network flow and local search, to solve these problems [3, 15, 21, 13, 12, 9].

Kleinberg and Tardos [19] formalized the metric labeling problem and its connections to Markov random fields and other classification problems (see [19] for a thorough description of the various connections). Metric

[1]This assumption is based on the connection of the labeling problem to the theory of Markov Random Fields (MRFs). In this theory the observed data or labeling of the objects is assumed to be obtained from a true labeling by a perturbation of independent random noise. The idea is to decide the most probable labeling given the observed data. A MRF can be defined by a graph on the objects with edges indicating dependencies between objects. Under the assumption that the probability that an object's label depends only on its neighbours labels, and if the MRF satisfies two standard assumptions of homogeneity and pairwise interactions, the labeling problem can be restated as the problem of finding a labeling f that maximizes the *a posteriori* probability $\mathbf{Pr}[f|f']$ where f' is the observed labeling. We refer the reader to [19] for more details on the connection of metric labeling to MRFs.

labeling also has rich connections to some well known problems in combinatorial optimization. It is related to the quadratic assignment problem, an extensively studied problem in Operations Research. A special case of metric labeling is the 0-extension problem, studied by Karzanov [17, 18]. There are no assignment costs in this problem, however, the graph contains a set of terminals, $t_1, \ldots, t_k$, where the label of terminal t_i is fixed in advance to i and the non-terminals are free to be assigned to any of the labels. Clearly, the 0-extension problem generalizes in a strong way the well-studied multi-way cut problem [10, 6, 11]. Karzanov [17, 18] showed that certain special cases (special metrics) of the 0-extension problem can be solved optimally in polynomial time.

Kleinberg and Tardos [19] obtained an $O(\log k \log \log k)$ approximation for the general metric labeling problem, where k denotes the number of labels in L and a 2-approximation for the uniform metric. However, the best lower bound on approximability for this problem is only max-snp hardness which follows from the max-snp hardness of the multi-way cut problem. Given the rich connections of this problem to other well-studied optimization problems and a variety of applications, a natural interesting question is to determine the approximability of the general problem as well as important special cases that arise in practice.

The Truncated Linear Metric There are many special cases of the problem that are interesting in their own right from both theoretical and applications point of view. One reason to solve special cases is because they give insight into the problem. In [19] the algorithm for the uniform case was used to approximate the general case. Very recently, Gupta and Tardos [14] considered the truncated linear metric, a special case motivated by its direct applications to two problems in computer vision, namely *image restoration* and *visual correspondence*. We briefly describe the application of this metric to the image restoration problem discussed earlier, see [14] for more details. Consider the case of grey scale images where the intensities are integer values. Our earlier description assumed that intensities of neighbouring pixels should be close since the image is typically smooth. This motivates a linear-like metric on the labels where the distance between two intensities i and j is $|i - j|$. However the assumption of closeness breaks down at object boundaries (here we are referring to objects in the image) where sharp changes in intensities happen. Thus, for the metric to be robust, neighbouring pixels that are actually at object boundaries (and hence differ in their intensities by large amounts) should not be penalized by arbitrarily large quantities. This motivated Gupta and Tardos to consider a truncated linear metric, where the distance between two intensities i and j is given by $d(i, j) = \min(M, |i - j|)$. Thus, the penalty is truncated at M; this is a natural (and non-uniform) metric for the problem at hand. A very similar reasoning applies to the visual correspondence problem, where the objective is to compare two images of the same scene for disparity. The labels here correspond to depth of the point in the image from the camera. For the truncated linear metric a 4-approximation was developed in [14] using a local search algorithm. The local moves in their algorithm make use of the flow network developed in [3, 15] that shows that the linear metric case can be solved to optimality.

For the image restoration application other distance functions have also been studied. In particular the quadratic distance function $d(i, j) = |i - j|^2$ and its truncated version $d(i, j) = \min\{M, |i - j|^2\}$ have been considered (see [20] and [16]). Unfortunately both these functions do not form a metric, however as we discuss shortly, we are able to provide non-trivial approximation ratios for them.

Our Results In this paper we address the problem of obtaining improved approximation guarantees for the metric labeling problem. Kleinberg and Tardos in their work [19] point out the difficulty of the general case as having to do with the absence of a "natural" IP formulation for the problem. They give an IP formulation for tree metrics and use Bartal's probabilistic tree approximations [1, 2] to reduce the general metric to a tree metric. In this work we give a natural IP formulation for the general problem. An advantage of our formulation is that it is applicable even to distance functions that are not metrics, for example the the quadratic distance function mentioned above. We substantiate the strength of this formulation by deriving both known results and new results using its linear relaxation. In particular, we show the following results on the integrality gap of our formulation.

- $O(\log k \log \log k)$ for general metrics and a factor 2 for the uniform metric.
- 1 for the linear metric and distances on the line defined by convex functions (not necessarily metrics).

- $2 + \sqrt{2} \simeq 3.414$ for the truncated linear metric.
- $O(\sqrt{M})$ for the truncated quadratic distance function.

The last result improves upon a $2M$ approximation known earlier. The integrality gaps we show for our formulation either match or improve upon the best previous approximation ratios for most of the cases known to us (in addition to the above we can also show that if G is a tree the integrality gap is 1). Our formulation allows us to present these results in a unified fashion. In the process we also improve the 4 approximation of Gupta and Tardos [14] for the important case of the truncated linear metric. We also show a reduction from the case with arbitrary assignment costs $c(u, i)$ to the case where $c(u, i) \in \{0, \infty\}$ for all u and i. The reduction preserves the graph G and the optimal solution, but increases the size of the label space from k labels to nk labels. Though this is undesirable in practice, it is useful from a theoretical point of view. In particular, this result implies that in order to find an $O(1)$-approximation for the general case, it suffices to focus attention on restricted instances with no assignment costs, but where each vertex is required to be assigned a label from its own set of labels. We believe our results and techniques are a positive step towards obtaining improved bounds for the general as well as interesting special cases of the metric labeling problem.

Independent of our work, Calinescu, Karloff, and Rabani [7] considered approximation algorithms for the 0-extension problem. They consider a linear programming formulation (which they call the *metric relaxation*) where they associate a length function with every edge of the graph and require that: (i) the distance between terminals t_i and t_j, for $1 \leq i, j \leq k$, is at least $d(i, j)$, and (ii) the length function is a semi-metric. We note that their formulation does not apply to the general metric labeling problem. They obtain an $O(\log k)$-approximation algorithm for the 0-extension problem using this formulation (and $O(1)$-approximation for planar graphs). Our linear programming formulation (when specialized to the 0-extension problem) induces a feasible solution for the metric relaxation formulation, by defining the length of an edge to be its transshipment cost (see Section 2). It is not hard to verify that this length function is a semi-metric. Calinescu, Karloff, and Rabani [7] also show a gap of $\Omega(\sqrt{\log k})$ on the integrality ratio of their formulation. It appears that their lower bound proof does not carry over (in any straight forward way) to our formulation. We note that the metric relaxation formulation optimizes over the set of all semi-metrics, while our formulation optimizes only over a subset of the semi-metrics. Whether our formulation is strictly stronger than the metric relaxation of [7] is an interesting open problem.

Outline: Section 2 describes our linear programming formulation for the general metric labeling problem. In Section 3, we analyze our formulation for uniform and linear metrics. Building on our rounding scheme for linear metric, we design and analyze a rounding procedure for the truncated linear metric in Section 4. In Section 5, we study the general metric labeling problem and show that the integrality gap of our formulation is bounded by $O(\log k \log \log k)$. We also describe here a transformation that essentially eliminates the role of the label cost assignment function.

2 The Linear Programming Formulation

We present a new linear integer programming formulation of the metric labeling problem. Let $x(u, i)$ be a $\{0, 1\}$-variable indicating that vertex u is labeled i. Let $x(u, i, v, j)$ be a $\{0, 1\}$-variable indicating that for edge $(u, v) \in E$, vertex u is labeled i and vertex v is labeled j.

Constraints (1) simply express that each vertex has to receive some label. Constraints (2) force consistency in the edge variables: if $x(u, i) = 1$ and $x(v, j) = 1$ they force $x(u, i, v, j)$ to be 1. Constraints (3) express the fact that (u, i, v, j) and (v, j, u, i) refer to the same edge; the redundancy helps in notation. We obtain a linear relaxation of the above program by allowing the variables $x(u, i)$ and $x(u, i, v, j)$ to take any non-negative value. We note that equality in (2) is important for the linear relaxation.

$$\text{(I)} \quad \min \sum_{u \in V} \sum_{i=1}^{k} c(u,i) \cdot x(u,i) \; + \sum_{(u,v) \in E} \sum_{i=1}^{k} \sum_{j=1}^{k} w(u,v) \cdot d(i,j) \cdot x(u,i,v,j)$$

subject to:

$$\sum_{i=1}^{k} x(u,i) = 1 \qquad \forall\, v \in V \tag{1}$$

$$\sum_{j=1}^{k} x(u,i,v,j) - x(u,i) = 0 \quad \forall\, u \in V, v \in N(u), \text{ and } i \in 1,\ldots,k \tag{2}$$

$$x(u,i,v,j) - x(v,j,u,i) = 0 \quad \forall\, u,v \in V, \text{ and labels } i,j \in 1,\ldots,k \tag{3}$$

$$x(u,i) \in \{0,1\} \qquad \forall\, u \in V \text{ and } i \in 1,\ldots,k \tag{4}$$

$$x(u,i,v,j) \in \{0,1\} \qquad \forall\, (u,v) \in E \text{ and } i,j \in 1,\ldots,k \tag{5}$$

With each edge $(u,v) \in E$ we associate a complete bipartite graph $H(u,v)$ The vertices of $H(u,v)$ are $\{u_1,\ldots,u_k\}$ and $\{v_1,\ldots,v_k\}$, i.e., they represent all possible labelings of u and v. There is an edge (u_i,v_j) connecting the pair of vertices u_i and v_j, $1 \le i,j \le k$. In the sequel, we refer to the edges of $H(u,v)$ as *links* to distinguish them from the edges of G. Suppose that the value of the variables $x(u,i)$ for all u and i has been determined. For an edge $(u,v) \in E$, we can interpret the variables $x(u,i,v,j)$ from a flow perspective. The contribution of edge $(u,v) \in E$ to the objective function of the linear program is the cost of the optimal transshipment of flow between $\{u_1,\ldots,u_k\}$ and $\{v_1,\ldots,v_k\}$, where: (i) the supply of u_i is $x(u,i)$ and the demand of v_j is $x(v,j)$ for $0 \le i,j \le k$, (ii) the cost of shipping a unit flow from u_i to v_j is $d(i,j)$. (The choice of the "supply" side and the "demand" side is arbitrary.) For the rest of the paper the quantity $d_{LP}(u,v)$ refers to the LP distance between u and v and is defined to be $\sum_{i,j} d(i,j) \cdot x(u,i,v,j)$.

3 Uniform Metrics and Linear Metrics

We now analyze the performance of our linear programming formulation on two natural special cases, namely, uniform metrics and linear metrics. Kleinberg and Tardos [19] showed a 2-approximation for the uniform metric case. Their approach is based on rounding the solution of a linear program formulated specifically for uniform metrics. We will show that our linear programming formulation dominates the one in [19], and thus also has an integrality gap of at most 2. For the case of linear metrics, Boykov *et al* [3], and Ishikawa and Geiger [15], have obtained exact algorithms, by reducing the problem to a minimum $\{s,t\}$-cut computation. We show here that, on linear metrics, our linear programming formulation gives an exact algorithm as well. Our analysis for the linear metric case plays an important role in the algorithm for the truncated linear metric case.

3.1 The Uniform Metric Case

Kleinberg and Tardos [19] formulated a linear program for the uniform metric and suggested the following iterative algorithm for rounding a solution to linear program (KT) below. Initially, no vertex is labeled. Each iteration consists of the following steps: (i) choose a label uniformly at random from $1,\ldots,k$ (say i), (ii) choose a real threshold θ uniformly at random from $[0,1]$, (iii) for all unlabeled vertices $u \in V$, u is labeled i if $\theta \le x(u,i)$. The algorithm terminates when all vertices are labeled. Kleinberg and Tardos [19] showed that the expected cost of a labeling obtained by this algorithm is a 2-approximation.

$$\text{(KT)}\quad \min \quad \sum_{v\in V}\sum_{i=1}^{k} c(u,i)\cdot x(u,i) \;+\; \frac{1}{2}\cdot \sum_{(u,v)\in E}\sum_{i=1}^{k} w(u,v)\cdot |x(u,i)-x(v,i)|$$

subject to:

$$\sum_{i=1}^{k} x(u,i) = 1 \qquad \forall\, u \in V$$

$$x(u,i) \geq 0 \quad \forall\, u \in V \text{ and } i \in 1,\ldots,k$$

We show that applying the rounding algorithm to an optimal solution obtained from linear program (I), yields the same approximation factor. Let $\bar{x}$ be a solution to (I). Note that for both (I) and (KT) the variables $x(u,i)$ completely determine the cost. We will show that cost of (KT) on $\bar{x}$ is smaller than that of (I). Both linear programs (I) and (KT) coincide regarding the labeling cost. Consider edge $(u,v) \in E$. We show that the contribution of (u,v) to the objective function of (I) is at least as high as the contribution to the objective function of (KT).

$$\begin{aligned}
\frac{1}{2}\cdot\sum_{i=1}^{k}|x(u,i)-x(v,i)| &= \frac{1}{2}\cdot\sum_{i=1}^{k}\left|\sum_{j=1}^{k}x(u,i,v,j)-\sum_{j=1}^{k}x(v,i,u,j)\right| \\
&\leq \frac{1}{2}\cdot\sum_{i=1}^{k}\left|\sum_{j=1,j\neq i}^{k}x(u,i,v,j)+\sum_{j=1,j\neq i}^{k}x(v,i,u,j)\right| \\
&= \frac{1}{2}\cdot\sum_{i=1}^{k}\sum_{j=1,j\neq i}^{k}2\cdot x(u,i,v,j) \qquad \text{since } x(u,i,v,j)=x(v,i,u,j) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k}d(i,j)\cdot x(u,i,v,j) \qquad \text{since } d(i,i)=0 \text{ and } d(i,j)=1, i\neq j
\end{aligned}$$

The last term is the contribution of (u,v) to the objective function of (I). We note that the example used in [19] to show that the integrality gap of (KT) is at least $2-1/k$, can be used to show the same gap for (I) as well.

3.2 The Line Metric Case

We now turn to the case of a linear metric and show that the value of an integral optimal solution is equal to the value of a fractional optimal solution of the LP (I). Without loss of generality, we can assume that the labels of the metric are integers $1,2,\ldots,k$. If the label set contains non-consecutive integers, we can add all the "missing" intermediate integers to the label set and set the cost of assigning them to every vertex to be infinite. In fact the linear program and the rounding can be generalized to the case where the labels are arbitrary points on the real line without any difficulty.

The rounding procedure: Let $\bar{x}$ be an optimal fractional solution to the linear program. We round the fractional solution as follows. Let θ be a real threshold chosen uniformly at random from $[0,1]$. For all i, $1 \leq i \leq k$, let

$$\alpha(u,i) = \sum_{j=1}^{i} x(u,j).$$

Each vertex $u \in V$ is labeled by the unique label i that satisfies $\alpha(u, i-1) < \theta \leq \alpha(u, i)$. Clearly all vertices are labeled since $\alpha(u, k) = 1$.

Analysis: We analyze the expected cost of the assignment produced by the rounding procedure. For each vertex $u \in V$, let $L(u)$ be a random variable that assumes the different labels that u can be assigned by the rounding procedure. It can be readily verified that the probability that $L(u) = i$ is equal to $x(u, i)$. This means that the expected labeling cost of vertex v is equal to $\sum_{i=1}^{k} x(u, i) \cdot c(u, i)$ which is precisely the assignment cost of u in the linear program (with respect to solution $\bar{x}$). We now fix our attention on the expected cost of the edges.

Lemma 3.1 *Consider edge* $(u, v) \in E$. *Then,*

$$\mathbf{E}\left[d((L(u), L(v))\right] = \sum_{i=1}^{k} |\alpha(u, i) - \alpha(v, i)|.$$

Proof. For the sake of the analysis we define auxiliary binary random variable $Z_1, \ldots, Z_{k-1}$ as follows. The variable Z_i is 1 if $\min\{L(u), L(v)\} \leq i$ and $\max\{L(u), L(v)\} > i$, and is 0 otherwise. In other words Z_i is 1 if i is in the interval defined by $L(u)$ and $L(v)$. It is easy to see that

$$d(L(u), L(v)) = \sum_{i=1}^{k-1} Z_i.$$

Therefore $\mathbf{E}\left[d(L(u), L(v))\right] = \sum_{i=1}^{k-1} \mathbf{E}\left[Z_i\right]$. We claim that

$$\mathbf{E}\left[Z_i\right] = \mathbf{Pr}\left[Z_i = 1\right] = |\alpha(u, i) - \alpha(v, i)|.$$

The lemma easily follows from this claim. To prove the claim, assume w.l.o.g. that $\alpha(u, i) \geq \alpha(v, i)$. If $\theta < \alpha(v, i)$ it is clear that $L(u), L(v) \leq i$ and if $\theta > \alpha(u, i)$ then $L(u), L(v) > i$: in both cases $Z_i = 0$. If $\theta \in (\alpha(v, i), \alpha(u, i)]$ then $L(u) \leq i$ and $L(v) > i$ which implies that $Z_i = 1$. Therefore $\mathbf{Pr}\left[Z_i = 1\right]$ is exactly $|\alpha(u, i) - \alpha(v, i)|$. □

We now estimate the contribution of an edge $(u, v) \in E$ to the objective function of the linear program. As indicated in Section 2, the contribution is equal to the cost of the optimal transshipment cost of flow in the complete bipartite graph $H(u, v)$ between $\{u_1, \ldots, u_k\}$ and $\{v_1, \ldots, v_k\}$, where the supply of u_i is $x(u, i)$ and the demand of v_j is $x(v, j)$ for $1 \leq i, j \leq k$. Recall that $d_{LP}(u, v) = \sum_{i,j} d(i, j) \cdot x(u, i, v, j)$.

Lemma 3.2 *For the line metric*

$$d_{LP}(u, v) \geq \sum_{i=1}^{k} |\alpha(u, i) - \alpha(v, i)|.$$

Proof. The crucial observation in the case of a linear metric is that flow can be "uncrossed". Let $i \leq i'$ and $j \leq j'$. Suppose that ε amount of flow is sent from u_i to $v_{j'}$ and from $u_{i'}$ to v_j. Then, uncrossing the flow, i.e., sending ε amount of flow from u_i to v_j and from $u_{i'}$ to $v_{j'}$ will not increase the cost of the transshipment. This means that the amount of flow sent from $\{u_1, \ldots, u_i\}$ to $\{v_1, \ldots, v_i\}$, for all i, $1 \leq i \leq k$, is precisely $\min\{\alpha(u, i), \alpha(v, i)\}$. Therefore, $|\alpha(v, i) - \alpha(u, i)|$ amount of flow is sent "outside" the label set $1, 2, \ldots, i$ and can be charged one unit of cost (with respect to i). Applying this argument to all values of i, $1 \leq i \leq k$, we get that the cost of the *optimal* transshipment of flow is precisely $\sum_{i=1}^{k} |\alpha(u, i) - \alpha(v, i)|$. The lemma follows. □

Hence, together with Lemma 3.1, we get that the expected cost of edge $(u, v) \in E$ after the rounding is no more than its contribution to the objective function of the linear program.

Theorem 1 *The integrality gap of the LP (I) for the line metric is* 1.

3.3 Distance functions on the Line defined by Convex functions

We now consider distance functions on the labels $1, \ldots, k$ on the integer line defined by strictly convex functions, that is $d(i,j) = f(|i-j|)$ where f is convex and increasing. Notice that such distance functions do not satisfy the metric property, in fact d is a metric if and only if f is concave and increasing. Our motivation for studying these metrics comes from the quadratic function $d(i,j) = |i-j|^2$ which is of particular interest in the image restoration application [16, 20] described earlier. We can show the following.

Theorem 2 *For any distance function on the line defined by a convex function, the integrality gap of our LP formulation is* 1.

The proof is based on the following facts, the proofs of which we omit from this version. Consider a feasible solution $\bar{x}$ to the LP. For any edge (u,v), if f is convex, the optimal cost transshipment flow in $H(u,v)$ is *non-crossing*. Further, for the rounding that we described, if the flow is non-crossing, **Pr** $[L(u) = i$ and $L(v) = j] = x(u,i,v,j)$. The theorem follows from this last fact trivially. Ishikawa and Geiger [16] show that the flow graph construction for the line metric can be extended for convex functions to obtain an optimal solution. The advantage of our approach is that the solution is obtained from a general formulation. This allows us to extend the ideas to obtain the first non-trivial approximation for the truncated quadratic distance function.

4 Improved Approximation for the Truncated Line Metric

In this section we use our LP formulation to give a $(2+\sqrt{2})$-approximation algorithm for the truncated line metric case. This improves the 4-approximation provided in [14]. We prefer to view the metric graph as a line with the truncation to M being implicit. This allows us to use, with appropriate modifications, ideas from Section 3.2. We round the fractional solution $\bar{x}$ once again using only the variables $x(u,i)$. Let $M' \geq M$ be an integer parameter that we will fix later. We repeat the following iteration until all vertices are assigned a label.

- Pick an integer ℓ uniformly at random in $[-M'+2, k]$. Let I_ℓ be the interval $[\ell, \ell+M'-1]$.
- Pick a real threshold θ uniformly at random from $[0,1]$.
- Let u be an unassigned vertex. If there is a label $i \in I_\ell$ such that

  $$\sum_{j=\ell}^{i-1} x(u,j) < \theta \leq \sum_{j=\ell}^{i} x(u,j),$$

  we assign i to u. Otherwise u is unassigned in this iteration.

The above algorithm generalizes the rounding algorithms for the uniform and line metrics in a natural way. Once the index ℓ is picked, the rounding is similar to that of the line metric in the interval I_ℓ. The difference is that a vertex might not get a label in an iteration. If two vertices u and v get separated in an iteration, that is only one of them gets assigned, our analysis will assume that their distance is M. This is similar to the analysis for the uniform metric case. Our improvement comes from a careful analysis of the algorithm which treats links differently based on whether their distance is linear or truncated. The analysis guides the choice of M' to obtain the best guarantee.

Let $L(u)$ and $L(v)$ be random variables that indicate the labels that get assigned to u and v by the algorithm.

Lemma 4.1 *In any given iteration, the probability of an unassigned vertex u getting a label i in that iteration is exactly $x(u,i) \cdot M'/(k+M'-1)$. The probability of u getting assigned in the iteration is $M'/(k+M'-1)$. Therefore* **Pr** $[L(u) = i] = x(u,i)$.

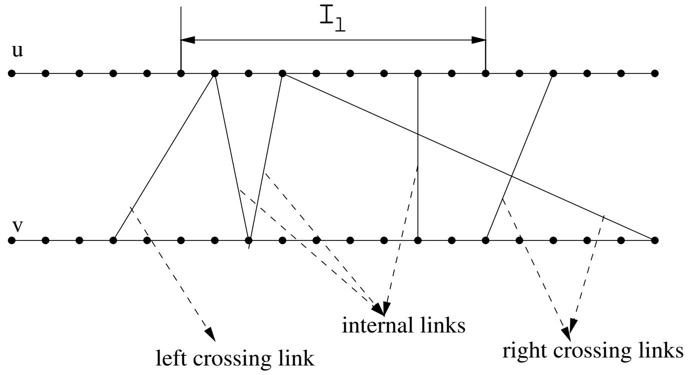

Figure 1: Illustration of interesting links for I_ℓ.

Proof. If ℓ is picked in the first step of an iteration, the probability of assigning i to u is exactly $x(u,i)$, if $i \in I_\ell$, and zero otherwise. The number of intervals that contain i is M', hence the probability of u getting i in an iteration is simply $x(u,i) \cdot M'/(k+M'-1)$. $\square$

It also follows from Lemma 4.1 that with high probability all vertices are assigned in $O(k \log n)$ iterations. The following lemma bounds the expected distance between $L(u)$ and $L(v)$ as a function of M'. Recall $d_{LP}(u,v) = \sum_{i,j} d(i,j) \cdot x(u,i,v,j)$.

Lemma 4.2 *The expected distance between $L(u)$ and $L(v)$ satisfies the following inequality.*

$$\mathbf{E}\left[d(L(u),L(v))\right] \leq (2 + \max\{\frac{2M}{M'}, \frac{M'}{M}\})d_{LP}(u,v).$$

Theorem 3 *There is a $(2+\sqrt{2})$-approximation algorithm for the metric labeling problem when the metric is truncated linear.*

Proof. We note that the algorithm and the analysis easily generalizes to the case when M' is a real number. We choose $M' = \sqrt{2}M$ and the theorem follows from Lemmas 4.1 and 4.2. $\square$

For the rest of the section we will restrict our attention to one particular edge $(u,v) \in E(G)$ and analyse the affect of the rounding on the expected distance with the goal of proving Lemma 4.2. To analyze the process we consider an iteration before which neither u and v are assigned. If in the current iteration only one of u or v is assigned a label, that is they are separated, we will pay a distance of M. If both of them are assigned then we pay a certain distance based on their labels in the interval. The main quantity of interest is the expected distance between the labels of u and v in a single iteration conditioned under the event that both of them are assigned in this iteration. Recall that a link refers to the edges in the complete bipartite graph $H(u,v)$.

Given an interval $I_\ell = [\ell, \ell+M'-1]$ we partition the interesting links for I_ℓ into three categories, *internal*, *left crossing*, and *right crossing*. The internal links denoted by $\text{INT}(I_\ell)$ are all the links (u,i,v,j) with $i,j \in I_\ell$, the left crossing links denoted by $\text{LCROSS}(I_\ell)$ are those with $\min\{i,j\} < \ell$ and and $\max\{i,j\} \in I_\ell$, and right crossing link denoted by $\text{RCROSS}(I_\ell)$ are those with $\min\{i,j\} \in I_\ell$ and and $\max\{i,j\} > \ell + M' - 1$. It is clear that no edge is both left and right crossing. Let $\text{CROSS}(I_\ell)$ denote $\text{LCROSS}(I_\ell) \cup \text{RCROSS}(I_\ell)$. See Figure 1 for an example. It is easy to see that e is not relevant to I_ℓ if $i,j \notin I_\ell$.

We set up some notation that we use for the rest of this section. For a link $e = (u,i,v,j)$ let $d_e = d(i,j)$ and $|e| = |i-j|$ be the truncated linear and linear distance respectively between i and j. The quantity x_e denotes $x(u,i,v,j)$. Consider a link e that crosses the interval I_ℓ and let i be the label of e that is internal to I_ℓ. We denote by e_ℓ the quantity $(\ell + M' - 1 - i)$, the linear distance from i to the right end of interval. The quantity $x(u, I_\ell)$ refers to $\sum_{i \in I_\ell} x(u,i)$, the flow of u assigned by the LP to labels in I_ℓ.

Lemma 4.3 *The probability of u and v being separated given that ℓ was chosen in the first step of the iteration is at most $\sum_{e \in \text{CROSS}(I_\ell)} x_e$.*

Proof. The probability of separation is exactly $|x(u, I_\ell) - x(v, I_\ell)|$ which can easily be seen to be upper bounded by $\sum_{e \in \text{CROSS}(I_\ell)} x_e$. □

Lemma 4.4 *For two vertices u and v unlabeled before an iteration let p_ℓ be the expected distance between them conditioned on the event that ℓ was chosen in the first step of the iteration and both were assigned a label in I_ℓ. Then*

$$p_\ell \leq \sum_{e \in \text{CROSS}(I_\ell)} e_\ell x_e + \sum_{e \in \text{INT}(I_\ell)} |e| x_e.$$

We provide some intuition before giving the formal proof. Once ℓ is fixed, the rounding is exactly the same as that for the line metric, restricted to the interval I_ℓ. From Lemma 3.1 we know the exact expected cost of the rounding. However we do not have an equivalent of Lemma 3.2 to bound the LP cost because I_ℓ is only a subinterval of the full line and also because of the truncation. The main difficulty is with links that belong to $\text{CROSS}(I_\ell)$. By charging each of the crossing links e, an amount equal e_ℓ (instead of d_e that LP paid), we are able to pay for the expected cost of rounding. In other words we charge the interesting links of I_ℓ to pay for the optimal linear metric transshipment flow induced by the fractional values $x(u,i)$ and $x(v,i)$, $i \in I_\ell$ when restricted to I_ℓ.

Proof. Fix an ℓ and w.l.o.g. assume that $x(u, I_\ell) \geq x(v, I_\ell)$. With probability $q = x(v, I_\ell)$ both u and v get labels from I_ℓ. We analyze the expected distance conditioned on this event. Once the interval I_ℓ is fixed the rounding is very similar to that of the linear metric case. For $0 \leq i < M'$ let $\alpha(u,i) = \sum_{j=\ell}^{j+i} x(u,j)$. The quantity $\alpha(u,i)$ sums the amount of flow of u in the first $i+1$ labels of the interval I_ℓ. In the following analysis we assume that distances within I_ℓ are linear and ignore truncation. This can only hurt us. Following the reasoning in Lemma 3.1 the expected distance between u and v is equal to $\sum_{i=0}^{M'-1} |\min\{q, \alpha(u,i)\} - \alpha(v,i)|$ which we upper bound by $\sum_{i=0}^{M'-1} |\alpha(u,i) - \alpha(v,i)|$. We claim that

$$\sum_{i=0}^{M'-1} |\alpha(u,i) - \alpha(v,i)| \quad \leq \quad \sum_{e \in \text{CROSS}(I_\ell)} e_\ell x_e + \sum_{e \in \text{INT}(I_\ell)} |e| x_e.$$

To prove this claim we consider each link $e \in \text{CROSS}(I_\ell) \cup \text{INT}(I_\ell)$ and sum its contribution to the terms $q_i = |\alpha(u,i) - \alpha(v,i)|$, $0 \leq i < M'$. Let $e = (u,a,v,b) \in \text{INT}(I_\ell)$. It is clear that e contributes exactly x_e to q_i if $a \leq i$ and $b > i$ or if $a > i$ and $b \leq i$. Otherwise its contribution is 0. Therefore the overall contribution of e to $\sum q_i$ is $x_e|a-b| = x_e|e|$.

Now suppose $e = (u,a,v,b) \in \text{LCROSS}(I_\ell)$. Assume w.l.o.g. that $a \geq \ell$ and $b < \ell$, the other case is similar. Link e will contribute x_e to $\alpha(u,i)$ for $a - \ell \leq i < M'$ and contributes 0 to $\alpha(v,i)$ for $0 \leq i < M'$ since b is outside the interval I_ℓ. Therefore the contribution of e to q_i is x_e for $a - \ell \leq i < M'$ and 0 otherwise, the overall contribution of e to $\sum q_i$ is $x_e|\ell + M' - 1 - a| = e_\ell x_e$. A similar argument holds for the case when $e \in \text{RCROSS}(I_\ell)$. We omit the details. □

Proof of Lemma 4.2. For a given iteration before which neither u nor v has a label, let $\mathbf{Pr}[u \wedge v]$, $\mathbf{Pr}[u \oplus v]$, and $\mathbf{Pr}[u \vee v]$ denote the probabilities that u and v are both assigned, exactly one of them is assigned, and at least one of them is assigned respectively. We upper bound the quantity $\mathbf{E}[d(L(u), L(v))]$ as follows. If u and v

are separated in some iteration we upper bound their resulting distance by M. Using this simplification we have the following.

$$\mathbf{E}\left[d(L(u),L(v))\right] \leq \frac{\mathbf{Pr}\left[u\oplus v\right]\cdot M + \mathbf{Pr}\left[u\wedge v\right]\cdot \mathbf{E}\left[d(L(u),L(v))|u\wedge v\right]}{\mathbf{Pr}\left[u\vee v\right]}.$$

We upper bound the right hand side of the above equation as follows.

- We lower bound $\mathbf{Pr}\left[u\vee v\right]$ by $\mathbf{Pr}\left[u\right]$ which by Lemma 4.1 is equal to $M'/(k+M'-1)$.
- We upper bound $\mathbf{Pr}\left[u\oplus v\right]$ by $\frac{1}{k+M'-1}\sum_\ell \sum_{e\in \text{CROSS}(I_\ell)} x_e$ using Lemma 4.3.
- By the definition of p_ℓ in Lemma 4.4 we get the following.

$$\mathbf{Pr}\left[u\wedge v\right]\mathbf{E}\left[d(L(u),L(v))|u\wedge v\right] = \frac{1}{k+M'-1}\sum_\ell p_\ell.$$

Putting all these together and using Lemma 4.4 to bound p_ℓ we get

$$\begin{aligned}
\mathbf{E}\left[d(L(u),L(v))\right] &\leq \frac{1}{M'}\sum_\ell\left(\sum_{e\in\text{CROSS}(I_\ell)}(M+e_\ell)x_e + \sum_{e\in\text{INT}(I_\ell)}|e|x_e\right)\\
&\leq \frac{1}{M'}\sum_e x_e\left(\sum_{\text{CROSS}(I_\ell)\ni e}(M+e_\ell) + \sum_{\text{INT}(I_\ell)\ni e}|e|\right)\\
&= \sum_e x_e\bar{d}_e
\end{aligned}$$

where $\bar{d}_e = \frac{1}{M'}(\sum_{\text{CROSS}(I_\ell)\ni e}(M+e_\ell) + \sum_{\text{INT}(I_\ell)\ni e}|e|)$. Lemma 4.5 shows that $\bar{d}_e$ is upper bounded by $(2+\max\{2M/M', M'/M\})d_e$. This finishes the proof since $d_{LP}(u,v) = \sum_e x_e d_e$. □

Lemma 4.5

$$\bar{d}_e = \frac{1}{M'}\sum_{\text{CROSS}(I_\ell)\ni e}(M+e_\ell) + \sum_{\text{INT}(I_\ell)\ni e}|e| \leq (2+\max\{\frac{2M}{M'},\frac{M'}{M}\})d_e.$$

Proof. We evaluate $\bar{d}_e$ separately for three different types of edges based on their lengths. Recall that $M' \geq M$ and hence $d_e \leq M \leq M'$ for all links e. Let e correspond to the link (u,i,v,j) in $H(u,v)$ and w.l.o.g. assume that $i \leq j$, the other case is similar, recall that $|e| = |i-j|$.

- $|e| \geq M'$: In this case it is clear that e is not an internal edge for any interval. Also $d_e = M$. Therefore

$$\begin{aligned}
M'\bar{d}_e &= \sum_{\text{CROSS}(I_\ell)\ni e}(M+e_\ell)\\
&= \sum_{\text{LCROSS}(I_\ell)\ni e}(M+e_\ell) + \sum_{\text{RCROSS}(I_\ell)\ni e}(M+e_\ell)\\
&= \sum_{\ell>j-M'}^{j}(M+M'+\ell-1-i) + \sum_{\ell>i-M'}^{i}(M+M'+\ell-1-i)\\
&\leq M'(2M+M')\\
&= M'(2+M'/M)d_e.
\end{aligned}$$

- $|e| < M$: $d_e = |e|$.

$$\begin{aligned}
M'\bar{d}_e &= \sum_{\text{CROSS}(I_\ell)\ni e} (M + e_\ell) + \sum_{\text{INT}(I_\ell)\ni e} |e| \\
&= \sum_{\text{LCROSS}(I_\ell)\ni e} (M + e_\ell) + \sum_{\text{RCROSS}(I_\ell)\ni e} (M + e_\ell) + \sum_{\text{INT}(I_\ell)\ni e} |e| \\
&= \sum_{\ell>i}^{j} (M + M' + \ell - 1 - i) + \sum_{\ell>i-M'}^{j-M'} (M + M' + \ell - 1 - i) + \sum_{\ell>j-M'}^{i} |e| \\
&\leq (2M' + 2M - |e|)|e| \\
&\leq (2M' + 2M)|e| \\
&= M'(2 + 2M/M')d_e.
\end{aligned}$$

- $M \leq |e| < M'$: $d_e = M$.

$$\begin{aligned}
M'\bar{d}_e &= \sum_{\text{LCROSS}(I_\ell)\ni e} (M + e_\ell) + \sum_{\text{RCROSS}(I_\ell)\ni e} (M + e_\ell) + \sum_{\text{INT}(I_\ell)\ni e} |e| \\
&= \sum_{\ell>i}^{j} (M + M' + \ell - 1 - i) + \sum_{\ell>i-M'}^{j-M'} (M + M' + \ell - 1 - i) + \sum_{\ell>j-M'}^{i} |e| \\
&= (2M' + 2M - |e|)|e| \\
&\leq M'(2M + M') \\
&\leq M'(2 + M'/M)d_e.
\end{aligned}$$

□

4.1 The Truncated Quadratic Distance on the Line

Consider the label space $1, 2, \ldots, k$ on the line where the distance function $d(i, j) = \min\{M, |i - j|^2\}$. This is the truncated version of the quadratic distance. We note that the quadratic distance is not a metric. However as mentioned earlier this distance function arises in image processing applications. In Subsection 3.3 we showed that our LP formulation gives an optimal solution for the quadratic distance on the line. For the truncated version of this distance we can use the algorithm from Section 4. By choosing $M' = \sqrt{M}$ we can show the following theorem.

Theorem 4 *The integrality gap of our LP on the truncated quadratic distance is $O(\sqrt{M})$.*

The best approximation ratio for this problem that follows from prior work is $2M$.

5 General Metrics

5.1 Integrality Gap on General Metrics

We now show that the integrality gap of our LP formulation is $O(\log k \log \log k)$ on general metrics. This gives an alternative way of obtaining the result of Kleinberg-Tardos [19]. The latter algorithm uses the approach of first approximating the given metric probabilistically by a hierarchically well-separated tree (HST) metric [1] and then using an LP formulation to solve the problem on tree metrics. The Kleinberg-Tardos LP formulation has only an $O(1)$ integrality gap on HST metrics. Since any arbitrary k-point metric can be probabilistically approximated by an HST metric with an $O(\log k \log\log k)$ distortion [2], their result follows. In contrast, our approach is based on directly using our LP formulation on the given general metric. As a first step, we use the LP solution to identify a deterministic HST metric approximation of the given metric such that the cost of our fractional solution on this HST metric is at most $O(\log k \log\log k)$ times the LP cost on the original metric. This first step is done by using the following proposition from Charikar *et al* [8].

Proposition 1 *Let d be an arbitrary k-point metric and let α be a non-negative function defined over all pairs of points in the metric. Then d can be deterministically approximated by an HST metric d_T such that*

$$\sum_{i,j} \alpha(i,j) \cdot d_T(i,j) \leq O(\log k \log\log k) \sum_{i,j} \alpha(i,j) \cdot d(i,j).$$

Given an optimal solution to our LP formulation, we apply Proposition 1 with the weight function $\alpha(i,j) = \sum_{(u,v)\in E} w(u,v) \cdot x(u,i,v,j)$ for $1 \leq i,j \leq k$. Thus $\alpha(i,j)$ is the fractional weight of edges between i and j. Let d_T denote the resulting HST metric. Since we are changing only the metric on the labels and not the assignments provided by the LP, the fractional solution is a feasible solution for this new metric and has cost at most $O(\log k \log\log k) \cdot C^*$ where C^* is the optimal LP cost for the original metric. Thus if we can now round our fractional solution on d_T by introducing only a constant factor increase in the solution cost, we will obtain an $O(\log k \log\log k)$-approximation algorithm. We prove this by showing that on any tree metric, our LP formulation is at least as strong as the Kleinberg-Tardos LP formulation (for tree metrics). We omit the proof and state the theorem.

Theorem 5 *The integrality gap of LP (I) on a k-point metric is $O(\log k \log\log k)$.*

5.2 Reduction to Zero-Infinity Assignment Costs

We now describe a transformation for the general problem that essentially allows us to eliminate the label assignment cost function. This transformation reduces an instance with arbitrary label assignment cost function c to one where each label assignment cost is either 0 or ∞. We refer to an instance of this latter type as *zero-infinity* instance. Our transformation exactly preserves the cost of any feasible solution but in the process increases the number of labels by a factor of n. This provides some evidence that the label cost assignment function does not play a strong role in determining the approximability threshold of the metric labeling problem. In particular, existence of a constant factor approximation algorithm for zero-infinity instance would imply a constant factor approximation algorithm for general instances as well.

From an instance $I = \langle c, d, w, L, G(V,E)\rangle$ of the general problem, we create an instance $I' = \langle c', d', w, L', G(V,E)\rangle$ of the zero-infinity variant as follows. We define a new label set $L' = \{i_u \mid i \in L \text{ and } u \in V\}$ i.e. we make a copy of L for each vertex in G. The new label cost assignment function is given by $c'(u, i_v) = 0$ if $v = u$ and ∞ otherwise. Thus each vertex has its own copy of the original label set and any finite cost solution to I' would assign each vertex a label from its own private copy.

Let $W_u = \sum_{v\in N(u)} w(u,v)$ for any vertex $u \in V$. The new distance metric on L' is defined in terms of the original distance metric as well as the original label cost assignment function. For $i \neq j$ or $u \neq v$,

$$d'(i_u, j_v) = d(i,j) + \frac{c(u,i)}{W_u} + \frac{c(v,j)}{W_v},$$

and $d'(i_u, i_u) = 0$. It can be verified that d' is indeed a metric, and that any solution to instance I can be mapped to a solution to instance I', and vice versa, in a cost-preserving manner. We omit details.

Theorem 6 *If there exists a $f(n,k)$-approximation algorithm for the zero-infinity instances, there exists a $f(n,nk)$-approximation algorithm for general instances.*

Acknowledgments

We thank Olga Veksler and Mihalis Yannakakis for useful discussions.

References

[1] Y. Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science*, pages 184–193, October 1996.

[2] Y. Bartal. On approximating arbitrary metrics by tree metrics. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 161–168, May 1998.

[3] Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 648–655, 1998.

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. In *Proceedings of the 7th IEEE International Conference on Computer Vision*, pages 377–384, 1999.

[5] Y. Boykov, O. Veksler, and R. Zabih. A new algorithm for energy minimization with disconitnuities. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, 1999.

[6] G. Calinescu, H. Karloff, and Y. Rabani. Improved approximation algorithms for multiway cut. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pages 48–52, 1998.

[7] G. Calinescu, H. Karloff, and Y. Rabani. Approximation algorithms for the 0-extension problem, manuscript, 2000.

[8] M. Charikar, C. Chekuri, A. Goel, and S. Guha. Rounding via trees: deterministic approximation algorithms for group steiner trees and k-median. In *Proceedings of the 30th ACM Symposium on Theory of Computing*, pages 114–123, 1998.

[9] F. S. Cohen. *Modeling and Applications of Stochastic Processes*, chapter Markov random fields for image modeling and analysis. Kluwer, 1986.

[10] E. Dalhaus, D. Johnson, C. Papadimitriou, P. Seymour, and M. Yannakakis. The complexity of multi-terminal cuts. *SICOMP*, 23(4):864–94, 1994.

[11] D.Karger, P. Klein, C. Stein, M. Thorup, and N. Young. Rounding algorithms for a geometric embedding of multiway cut. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pages 668–678, 1999.

[12] R. Dubes and A. Jain. Random field models in image analysis. *J. of Applied Statistics*, 16, 1989.

[13] D. Greig, B. Porteous, and A. Seheult. Exact maximum *a posteriori* estimation for binary images. *J. of Royal Statistical Society B*, 48, 1989.

[14] A. Gupta and E. Tardos. Constant factor approximation algorithms for a class of classification problems. In *Proceedings of the 32nd Annual ACM Symposium on the Theory of Computing*, 2000. To appear.

[15] H. Ishikawa and D. Geiger. Segmentation by grouping junctions. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 125–131, 1998.

[16] H. Ishikawa and D. Geiger. Mapping Image Restoration to a Graph Problem. In *Proceedings of the 1999 IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, pages 20–23, 1999.

[17] A. Karzanov. Minimum 0-extension of graph metrics. *Europ. J. Combinat.*, 19:71–101, 1998.

[18] A. Karzanov. A combinatorial algorithm for the minimum $(2, r)$-metric problem and some generalizations. *Combinatorica*, 18(4):549–569, 1999.

[19] J. Kleinberg and E. Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and markov random fields. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 14–23, 1999.

[20] S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, 1995.

[21] S. Roy and I. Cox. A maximum-flow formulation of the N-camera stereo correspondence problem. In *Proceedings of the International Conference on Computer Vision*, pages 492–499, 1998.