Combinatorial Optimization in Computer Vision

Chapter 5: Multilabel Problems with Linearly Ordered Label Space and with Convex Pairwise Potential

WS 2011/12

Ulrich Schlickewei
Computer Vision and Pattern Recognition Group
Technische Universität München

Plan for Today

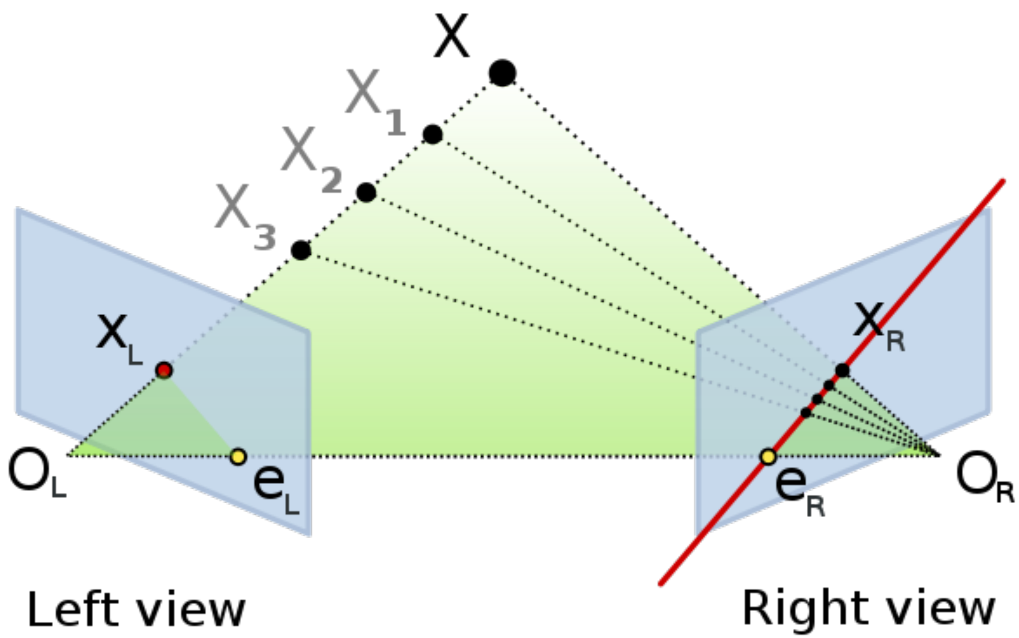
Reminder of Epipolar Geometry: The Stereo Problem

MRFs with Linearly Ordered Label Set and Convex Prior

Ishikawa‘s Construction

The Stereo Problem

Goal of Stereo Vision: Given two input images from two different perspectives, infer 3D coordinates of objects in the image.



Epipolar Plane

Given a pixel x_L in the left image, the epipolar plane is the plane spanned by the line between the camera centers and by the line between x_L and O_L. The main task in stereo vision is to determine the correct correspondence pixel x_R on the red line, the so-called epipolar line.

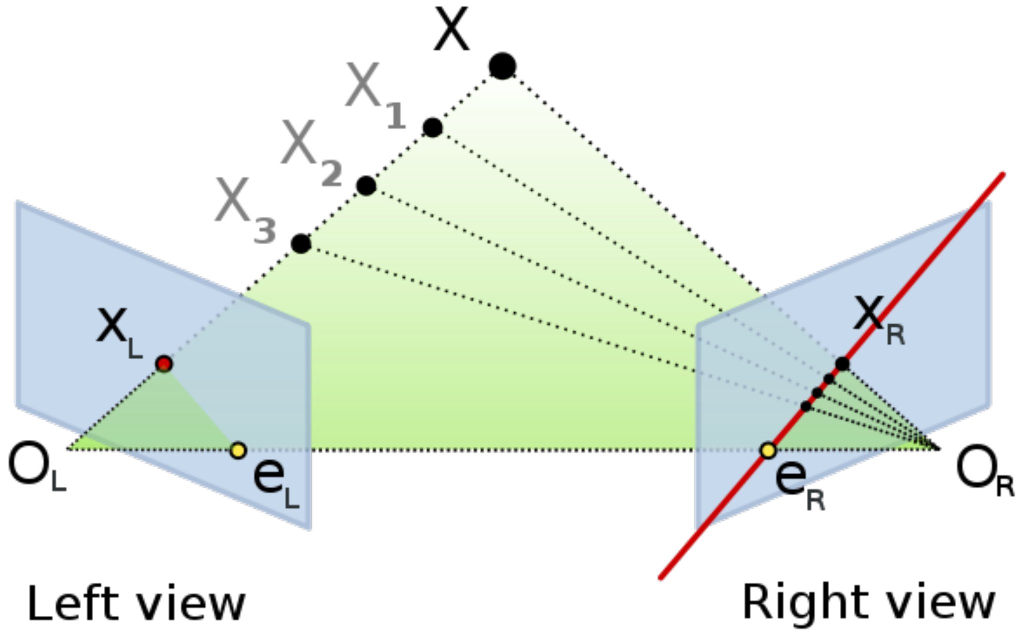
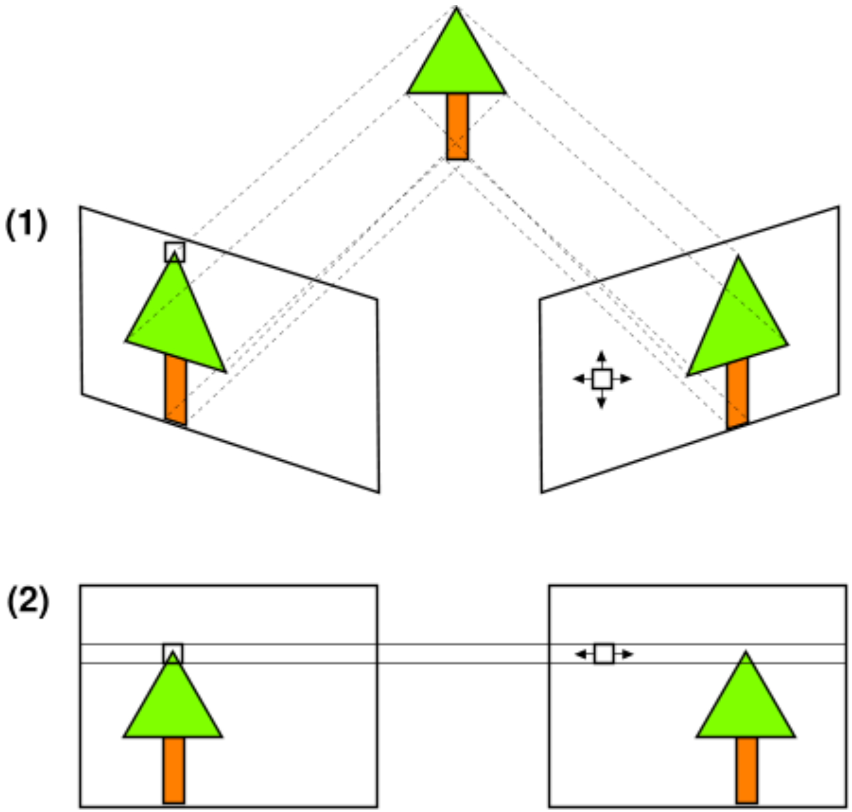


Image Rectification



We apply image rectification. After that preprocessing step which depends on a good calibration of the cameras, the epipolar lines are horizontal and we have to determine the pixel disparity

Determination of Disparity

The disparity $d : \Omega \to \mathbb{R}$ tells us, that for a given pixel in the left image $p_L = (x_L, y_L)$ the corresponding pixel in the right image is given by $p_R = (x_L + d(p_L), y_L)$.

In the variational setting the disparity is estimated by minimizing an energy similar to

$$\begin{aligned} E(d) &= E_{\text{data}}(d) + E_{\text{reg}}(d) \\ &= \int_\Omega |I_L(x_L, y_L) - I_R(x_L + d, y_L)|^2 + \int_\Omega |\nabla d|. \end{aligned}$$

Depth from Disparity

- The depth can be easily determined from the disparity if the camera parameters are known. It turns out that depth and disparity are inversely proportional.
- In the sequel we are going to learn an algorithm which can solve energy minimization problems of the above type globally in polynomial time.

Plan for Today

Reminder of Epipolar Geometry: The Stereo Problem

MRFs with Linearly Ordered Label Set and Convex Prior

Ishikawa‘s Construction

Multilabel Problems with Linearly Ordered Label Space

Consider a symmetric, directed graph $G=(V,E)$. In the following we study the inference problem for a Markov Random Field $X=(X_v)_{v\in V}$ on $G=(V,E)$.

Let

- $\mathcal{L}=\{1,\dots N\}$ be a linearly ordered set of label in which the MRF takes values
- $g:\mathbb{Z}\to\mathbb{R}$ be a convex and symmetric function, that is
  - $g''(z)=\{g(z+1)-g(z)\}-\{g(z)-g(z-1)\}\geq 0$ for all $z\in\mathbb{Z}$
  - $g(z)=g(-z)$ for all $z\in\mathbb{Z}$
- $h_v:\mathcal{L}\to\mathbb{R}$ be a function for each $v\in V$
- $\alpha_{uv}\geq 0$ be constants.

MRF Inference

Finally assume that the MRF X has the Gibbs potential

$$E(x) = \sum_{v \in V} h_v(x_v) + \sum_{v \sim w} \alpha_{uv} g(x_v - x_w).$$

Then, the MRF MAP-inference problem for X is the optimization problem

$$x^* = \operatorname{argmin}_{x \in \mathcal{L}^{|V|}} E(x).$$

We will now see, that due to the convexity of g this problem is tractable.

Plan for Today

Reminder of Epipolar Geometry: The Stereo Problem

MRFs with Linearly Ordered Label Set and Convex Prior

Ishikawa's Construction

Overview

Idea: Construct a network $N = (W, F, s, t, c)$ such that

a) There is a bijection

$$\{\text{cuts of } N\} \longleftrightarrow \{\text{configurations of MRF } X\}$$

b) The cost of a cut in N is the same as the energy of the corresponding MRF configuration (up to a constant).

Once, this network is constructed, the MRF inference problem is reduced to a min-cut problem which can be solved in polynomial time.

Definition of the Network

The network is defined by introducing

- **Vertices**

$$W = (V \times \mathcal{L}) \cup \{s,t\} = \{u_{v,l} \mid v \in V,\ l \in \mathcal{L}\} \cup \{s,t\}.$$

  Besides source and sink, there are indicator vertices for each possible label assignment.

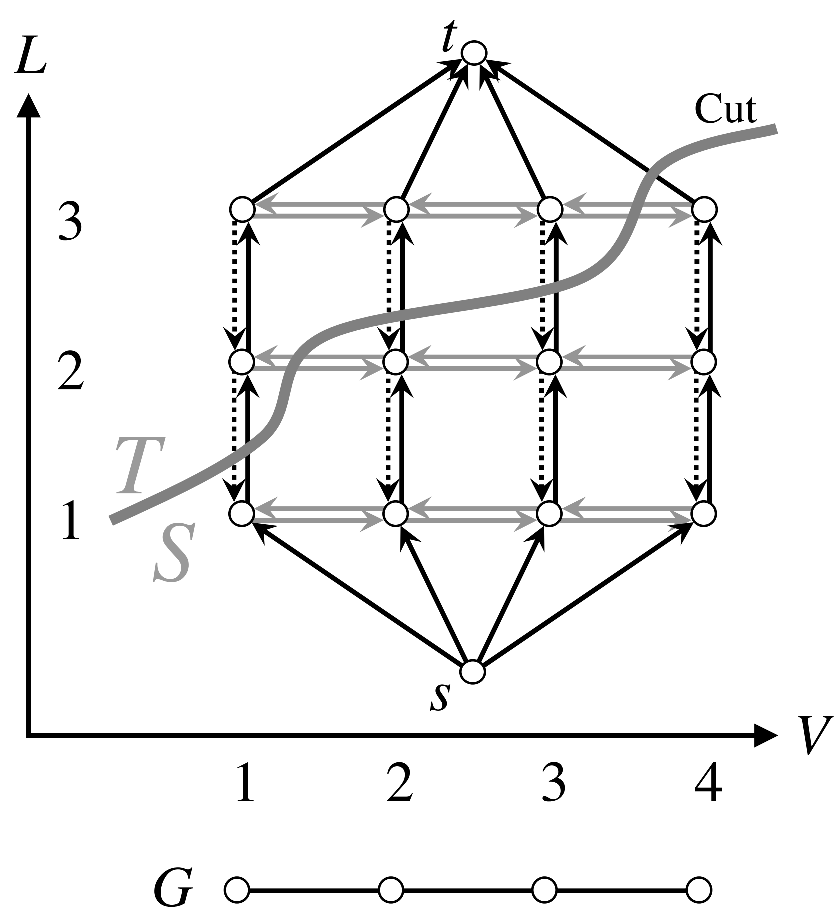
- **Edges**

$$F = F_d \cup F_c \cup F_p,$$

  where F_d is the set of data edges, F_c the set of constraint edges and F_p is the set of penalty edges (all defined below)

- **Capacities** defined together with the edges.

Indicator Vertices for Label Assignments



Data Edges

Data egdes connect the label vertices $u_{v,l}$ of one given „ground vertex“ $v \in V$ as follows:

- $u_{v,1}$ is connected with the source s
- $u_{v,l}$ is connected with $u_{v,l+1}$ if $1 \leq l \leq N-1$
- $u_{v,N}$ is connected with the sink t.

Thus, the data edges are given by

$$F_D = \cup_{v \in V} F_{D,v}$$

with

$$F_{D,v} = \{(s, u_{v,1}), (u_{v,1}, u_{v,2}), \ldots, (u_{v,N}, t)\}.$$

Capacities of Data Edges

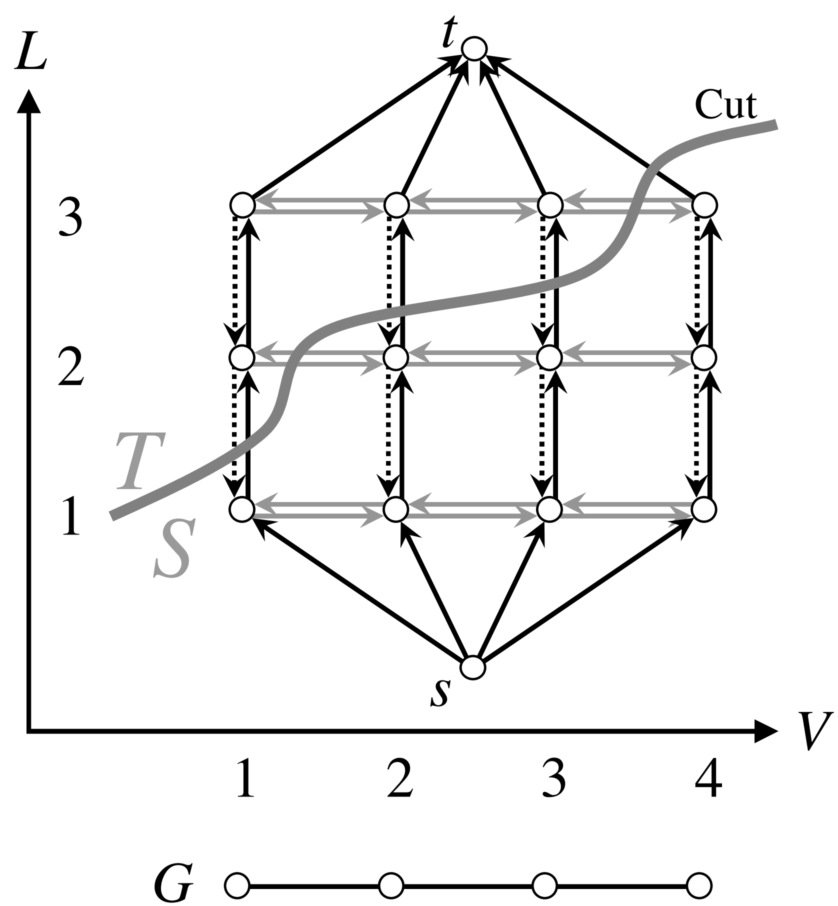
The capacities of the data edges corresponding to „ground vertex“ $v \in V$

$$F_{D,v} = \{(s, u_{v,1}), (u_{v,1}, u_{v,2}), \ldots, (u_{v,N}, t)\}.$$

are defined as follows:

- $c(s, u_{v,1}) = \infty$
- $c(u_{v,l}, u_{v,l+1}) = h_v(l)$
- $c(u_{v,N}, t) = h_v(N).$

Interpretation of a Cut in the Network I



Interpretation of a Cut in the Network II

Note that for each $v \in V$ at least one edge of the form $(u_{v,i}, u_{v,i+1})$ or the edge (u_N, t) is part of the cut (because otherwise the source is connected to the sink).

Convention: If edge $(u_{v,i}, u_{v,i+1})$ is part of the cut, we interpret the cut as assigning label i to vertex v. If the edge (u_N, t) is part of the cut, we interpret this as assigning label N to the vertex.

This convention only makes sense, if no more than one edge from $F_{D,v}$ is part of the cut. This will be ensured by the constraint edges.

Constraint Edges I

Constraint edges ensure that in every column only one edge belongs to a cut. They are defined as follows:

$$F_c = \cup_{v \in V} F_{c,v}$$

with

$$F_{c,v} = \{(u_{v,i+1}, u_{v,i}) \mid i = 1, \ldots N-1\}.$$

The capacity of constraint edges is set to infinity:

$$c(u_{v,i+1}, u_{v,i}) = \infty.$$

Constraint Edges II

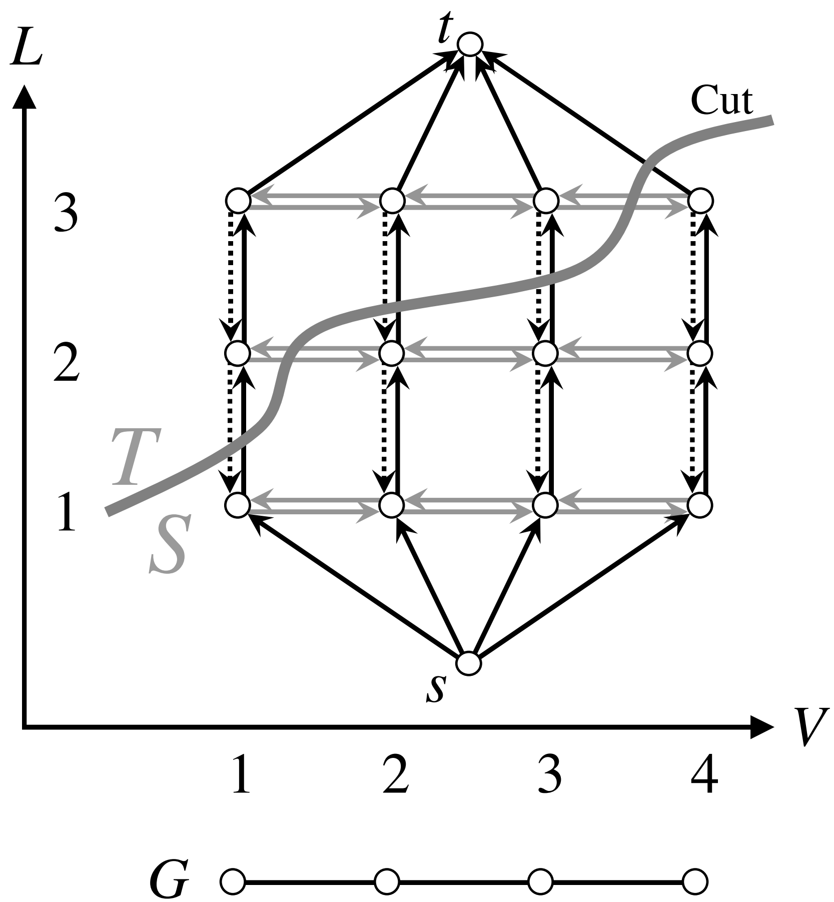
Lemma: A cut in the network $N = (W, F, s, t, c)$ involves more than one data edge from

$$F_{D,v} = \{(s, u_{v,1}), (u_{v,1}, u_{v,2}), \ldots, (u_{v,N}, t)\}$$

if and only if it involves a constraint edge.

The lemma shows that with the introduction of constraint edges we have ensured that only one label is assigned to each „ground vertex“ by a cut of finite cost.

Visualization of Constraint Edges



Penalty Edges I

Penalty edges are introduced for modelling the pairwise potential

$$\sum_{v\sim w} \alpha_{uv} g(x_v - x_w).$$

In the sequel we assume for simplicity that

- $\alpha_{uv} = 1$
- $g''(x) = \{f(x+1) - f(x)\} - \{f(x) - f(x-1)\} = 0 \text{ for } x \notin \pm\mathcal{L} \cup \{0\}$ (this is no loss of generality because these values of x do not appear in our optimization problem)

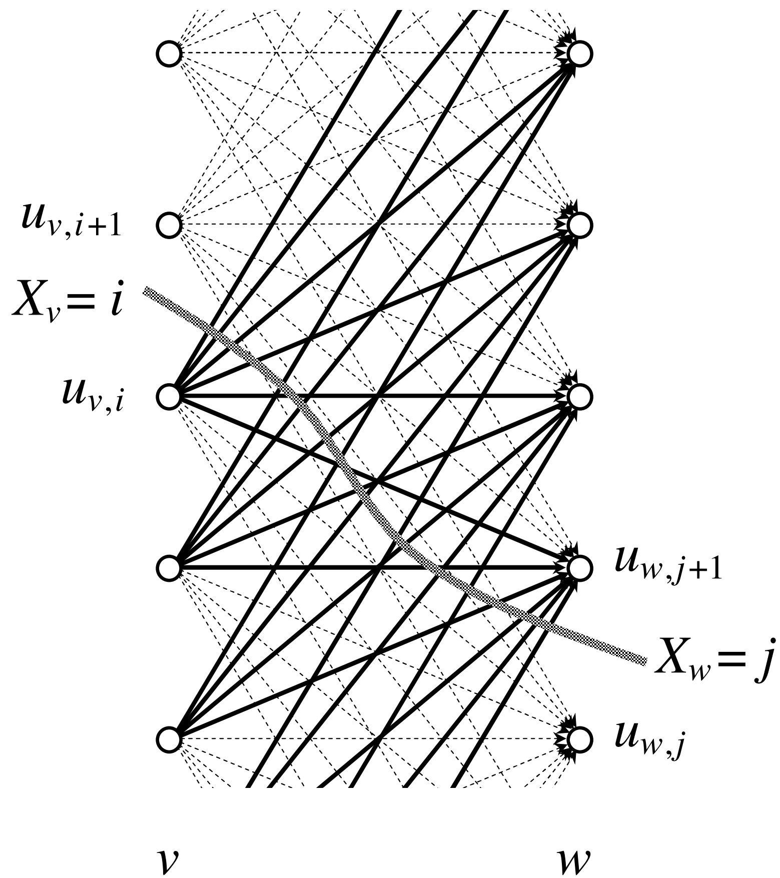
Penalty Edges II

Penalty edges connect label vertices corresponding to neighboring „ground vertices“:

$$F_p = \{(u_{v,i}, u_{w,j}) \mid (v,w) \in E,\ i,j \in \mathcal{L}\}.$$

To derive the correct capcities $c(u_{v,i}, u_{w,j})$, we first determine which penalty edges are involved in a cut corresponding to an assignment of labels.

Penalty Edges III



Penalty Edges IV

If $(u_{v,i}, u_{i+1})$ and $(u_{w,j}, u_{w,j+1})$ are in the cut for two neighboring „ground vertices" v, w, then

$$\left.\begin{array}{l} u_{v,l} \in T \text{ for } l > i \\ u_{w,k} \in S \text{ for } k \leq j \end{array}\right\} \quad (u_{v,l}, u_{w,k}) \text{ in the cut for } l > i, k \leq j$$

and similarly

$$\left.\begin{array}{l} u_{v,l} \in S \text{ for } l \leq i \\ u_{w,k} \in S \text{ for } k > j \end{array}\right\} \quad (u_{w,k}, u_{v,l}) \text{ in the cut for } l \leq i, k > j$$

Penalty Edges V

Therefore, the penalty edges lead to the following pairwise cost for the labelling $x_v = i, x_w = j$

$$f(i,j) = \sum_{l>i} \sum_{k \leq j} c(u_{v,l}, u_{w,k}) + \sum_{l \leq i} \sum_{k>j} c(u_{w,k}, u_{v,l}).$$

In order to model our original inference problem by cuts in the network, we need to have

$$g(i-j) = f(i,j) + \text{constant independent of } i,j$$

Penalty Edges VI

Lemma: If the function

$$\tilde{g}(i-j) := f(i,j)$$

is well-defined (that is, if $f(i,j)$ depends only on the difference $i-j$, then

$$\tilde{g}''(i-j) = c(u_{v,i}, u_{w,j}) + c(u_{w,j}, u_{v,i}).$$

In particular, since the capacities in our network are non-negative, $\tilde{g}$ is convex.

Plan: Define edge weights in such a way that $f(i,j)$ depends only on $i-j$, then show that in this case g and $\tilde{g}$ do only differ by a constant.

Penalty Edges VII

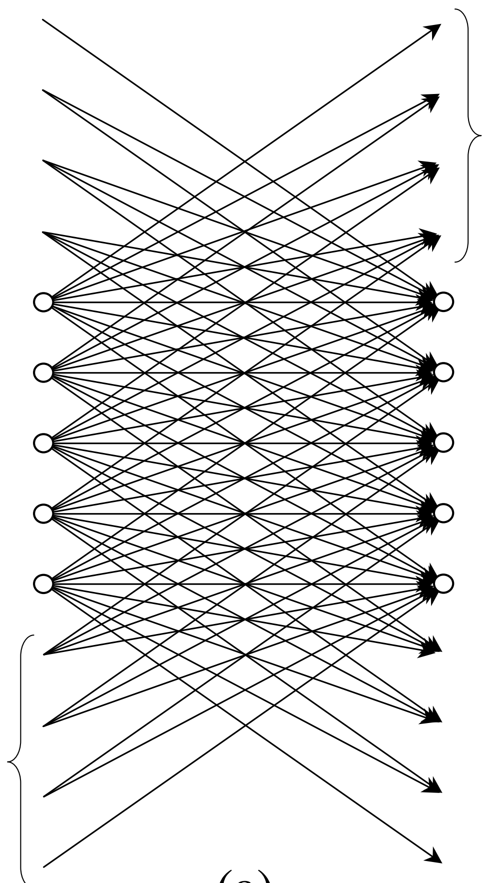
We define capacities for the penalty edges by

$$c(u_{v,i}, u_{w,j}) := \frac{1}{2} g''(i-j).$$

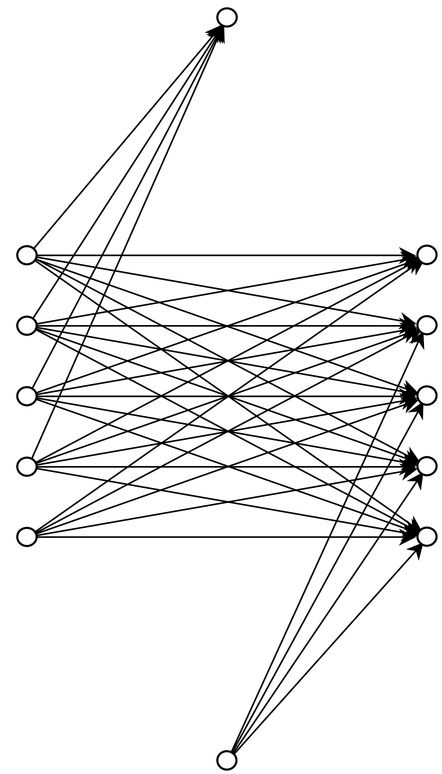
With this definition, the capacities do only differ from label differences. If we had an infinite label set, we could infer that $f(i,j)$ does only depend on $i-j$.

Instead of introducing additional vertices $u_{v,r}$ for $r \leq 0,\ r > N$ (as in depicted in (a) on the next slide), we add the correspoding capacities $c(u_{v,r}, u_{w,j}) = x''(r-j)$ to edges connecting w and s (resp. t) if $r \leq 0$ (resp. $r > N$) .

Penalty Edges VIII



(a)



(b)

Penalty Edges IX

Proposition: With this choice of capacities of the penalty egdes, the function $\tilde{g}(i-j) := f(i,j)$ is well-defined. Furthermore, it is a convex and symmetric function which satisfies

$$\tilde{g}''(x) = g''(x) \;\; \forall x \in \mathbb{Z}.$$

Therefore, $\tilde{g}$ and g only differ by a constant.

The proposition shows that with Ishikawa's graph construction our MRF-inference problem can be solved globally in polynomial time.