Full text available at: http://dx.doi.org/10.1561/0600000084

Discrete Graphical Models — An Optimization Perspective

Full text available at: http://dx.doi.org/10.1561/0600000084

Full text available at: http://dx.doi.org/10.1561/0600000084

Discrete Graphical Models — An Optimization Perspective

Bogdan Savchynskyy
Heidelberg University
bogdan.savchynskyy@iwr.uni-heidelberg.de

the essence of knowledge
Boston — Delft

Full text available at: http://dx.doi.org/10.1561/0600000084

Foundations and Trends® in Computer Graphics and Vision

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

B. Savchynskyy. *Discrete Graphical Models — An Optimization Perspective*. Foundations and Trends® in Computer Graphics and Vision, vol. 11, no. 3-4, pp. 160–429, 2019.

ISBN: 978-1-68083-639-4
© 2019 B. Savchynskyy

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Full text available at: http://dx.doi.org/10.1561/0600000084

Foundations and Trends® in Computer Graphics and Vision

Volume 11, Issue 3-4, 2019

Editorial Board

Editor-in-Chief

Brian Curless
University of Washington

William T. Freeman
MIT

Luc Van Gool
KU Leuven and ETH Zurich

Editors

Marc Alexa
TU Berlin

Kavita Bala
Cornel

Ronen Basri
Weizmann Institute of Science

Peter Belhumeur
Columbia University

Andrew Blake
Microsoft Research

Chris Bregler
Facebook-Oculus

Joachim Buhmann
ETH Zurich

Michael Cohen
Microsoft Research

Paul Debevec
USC Institute for Creative Technologies

Julie Dorsey
Yale

Fredo Durand
MIT

Olivier Faugeras
INRIA

Rob Fergus
NYU

Mike Gleicher
University of Wisconsin

Richard Hartley
Australian National University

Aaron Hertzmann
Adobe Research

Hugues Hoppe
Microsoft Research

C. Karen Liu
Georgia Tech

David Lowe
University of British Columbia

Jitendra Malik
Berkeley

Steve Marschner
Cornell

Shree Nayar
Columbia

James O'Brien
Berkeley

Tomas Pajdla
Czech Technical University

Pietro Perona
California Institute of Technology

Marc Pollefeys
ETH Zurich

Jean Ponce
Ecole Normale Superieure

Long Quan
HKUST

Cordelia Schmid
INRIA

Steve Seitz
University of Washington

Amnon Shashua
Hebrew University

Peter Shirley
University of Utah

Noah Snavely
Cornell

Stefano Soatto
UCLA

Richard Szeliski
Microsoft Research

Joachim Weickert
Saarland University

Song Chun Zhu
UCLA

Andrew Zisserman
Oxford

Full text available at: http://dx.doi.org/10.1561/0600000084

Editorial Scope

Topics

Foundations and Trends® in Computer Graphics and Vision publishes survey and tutorial articles in the following topics:

- Rendering
- Shape
- Mesh simplification
- Animation
- Sensors and sensing
- Image restoration and enhancement
- Segmentation and grouping
- Feature detection and selection
- Color processing
- Texture analysis and synthesis
- Illumination and reflectance modeling
- Shape representation
- Tracking
- Calibration
- Structure from motion
- Motion estimation and registration
- Stereo matching and reconstruction
- 3D reconstruction and image-based modeling
- Learning and statistical methods
- Appearance-based matching
- Object and scene recognition
- Face detection and recognition
- Activity and gesture recognition
- Image and video retrieval
- Video analysis and event recognition
- Medical image analysis
- Robot localization and navigation

Information for Librarians

Foundations and Trends® in Computer Graphics and Vision, 2019, Volume 11, 4 issues. ISSN paper version 1572-2740. ISSN online version 1572-2759. Also available as a combined paper and online subscription.

Full text available at: http://dx.doi.org/10.1561/0600000084

Contents

Full text available at: http://dx.doi.org/10.1561/0600000084

Full text available at: http://dx.doi.org/10.1561/0600000084

Full text available at: http://dx.doi.org/10.1561/0600000084

Discrete Graphical Models — An Optimization Perspective

Bogdan Savchynskyy

Heidelberg University; bogdan.savchynskyy@iwr.uni-heidelberg.de

ABSTRACT

This monograph is about combinatorial optimization. More precisely, about a special class of combinatorial problems known as *energy minimization* or *maximum a posteriori (MAP) inference in graphical models*, closely related to *weighted* and *valued constraint satisfaction problems* and having tight connections to *Markov random fields* and *quadratic pseudo-boolean optimization*. What distinguishes this monograph from a number of other monographs on graphical models is its focus: It considers graphical models, or, more precisely, MAP-inference for graphical models, purely as a combinatorial optimization problem. Modeling, applications, probabilistic interpretations and many other aspects are either ignored here or find their place in examples and remarks only.

Combinatorial optimization as a field is largely based on five fundamental topics: (i) integer linear programming and polyhedral optimization; (ii) totally unimodular matrices and the class of min-cost-flow problems; (iii) Lagrange decompositions and relaxations; (iv) dynamic programming and (v) submodularity, matroids and greedy algorithms. Each of these topics found its place in this monograph, although to a variable extent. The covering of each respective topic

Bogdan Savchynskyy (2019), “Discrete Graphical Models — An Optimization Perspective”, Foundations and Trends® in Computer Graphics and Vision: Vol. 11, No. 3-4, pp 160–429. DOI: 10.1561/0600000084.

Full text available at: http://dx.doi.org/10.1561/0600000084

reflects its importance for the considered MAP-inference problem.

Since optimization is the primary topic of this monograph, we mostly stick to the terminology widely used in optimization and where it was possible we tried to avoid the graphical models community-specific technical terms. The latter differ from sub-community to sub-community and, in our view, significantly complicate the information exchange between them.

The same holds also for the presentation of material in this monograph. If there is a choice when introducing mathematical constructs or proving statements, we prefer more general mathematical tools applicable in the whole field of operations research rather than to stick to graphical model-specific constructions. We additionally provide the graphical model-specific constructions if it turns out to be easier than the more general one. This way of presentation has two advantages. A reader familiar with a more general technique can grasp the new material faster. On the other hand, the monograph may serve as an introduction to combinatorial optimization for readers unfamiliar with this subject. To make the monograph even more suitable for both categories of readers we explicitly separate the mathematical optimization background chapters from those specific to graphical models.

We believe, therefore, that the monograph can be useful for undergraduate and graduate students studying optimization or graphical models, as well as for experts in optimization who want to have a look into graphical models. Moreover, we believe that even experts in graphical models can find new views on the known facts as well as a novel presentation of less known results in the monograph. These are for instance (i) a simple and general proof of equivalence of different acyclic Lagrange decompositions of a graphical model; (ii) a general scheme for analyzing convergence of

Full text available at: http://dx.doi.org/10.1561/0600000084

dual block-coordinate descent methods; (iii) a short and self-contained analysis of a linear programming relaxation for binary graphical models, its persistency properties and its relation to quadratic pseudo-boolean optimization.

The present monograph is based on lectures given to undergraduate students of Technical University of Dresden and Heidelberg University. The selection of material is done in a way that it may serve as a basis for a semester course.

Full text available at: http://dx.doi.org/10.1561/0600000084

Notation

To simplify reading of the monograph, some frequently used notations are collected here. Some of them, which we assume to be quite standard, are used without additional notice in the text. Others, typically more specialized, are introduced in the monograph. For those we point out the section and the page they are defined in.

Full text available at: http://dx.doi.org/10.1561/0600000084

Standard notation

$\mathbb{N}$	the set of natural numbers
$\mathbb{Z}$	the set of integer numbers
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^n$	an n-dimensional vector space over the field of real numbers
$\mathbb{R}^n_+$	the set of vectors with non-negative coordinates in $\mathbb{R}^n$, i.e. $\{x \in \mathbb{R}^n \colon x_i \geq 0,\ i = 1, 2, \dots, n\}$; for $n = 1$ the notation simplifies to $\mathbb{R}_+$
$x \in \mathcal{A}^{\mathcal{B}}$	For any set A and any finite set $\mathcal{B}$, this notation stands for a vector x with $\|\mathcal{B}\|$ coordinates indexed by elements of $\mathcal{B}$, for each $b \in \mathcal{B}$ it holds that $x_b \in \mathcal{A}$. The only exception from this rule is the notation $\Delta^{\mathcal{B}}$, see below.
$x \geq y$	comparison operations are applied coordinate-wise to vectors and point-wise to functions
$[\![\cdot]\!]$	denotes the Iverson brackets, that is, for any predicate A it holds that $[\![A]\!] = 1$ if A is true, otherwise $[\![A]\!] = 0$
$\langle c, x \rangle$	the inner product, i.e. $\langle c, x \rangle = \sum_{i=1}^n c_i x_i$
∇f	gradient of the function f
$O(\cdot)$	for two functions $f \colon \mathbb{N} \to \mathbb{N}$ and $g \colon \mathbb{N} \to \mathbb{N}$ one writes $f = O(g)$, if there is a constant $c > 0$ and a number $n_0 \in \mathbb{N}$ such that $f(n) \leq c \cdot g(n)$ for all $n \geq n_0$

Standard abbreviations

w.r.t.	with respect to
w.l.o.g.	without loss of generality
s.t.	subject to

Full text available at: http://dx.doi.org/10.1561/0600000084

Notation defined in the monograph

$\mathcal{N}_b(u)$	set of graph vertexes incident to vertex u, see §1.1, page 9
$\delta_{\mathcal{G}}(y)$, $\delta(y)$	binary representation of the labeling y, i.e. a binary vector with non-zero coordinates corresponding to the labels y_u, $u \in \mathcal{Y}_u$, and label pairs (y_u, y_v), $uv \in \mathcal{E}$, see page 49;
$\mathcal{I}$	the set of indexes of the cost vector of a graphical model; $\lvert\mathcal{I}\rvert$ is equal to the number of coordinates of the cost vector, see §1.1 on page 11
$\mathrm{vrtx}(P)$	set of vertexes of the polyhedron P, see Definition 3.19 on page 32
Δ^n	n-dimensional simplex, see Definition 3.21 on page 33
$\Delta^{\mathcal{X}}$	$\lvert\mathcal{X}\rvert$-dimensional simplex, with coordinates indexed by elements of $\mathcal{X}$, see Definition 3.21 on page 33
$\mathrm{conv}(X)$	convex hull of X, see Definition 3.28 on page 34
$\mathrm{mi}[\theta_w]$	binary vector with non-zero coordinates corresponding to locally minimal values of the cost vector θ_w, see page 95
$\mathrm{nz}[\mu]$	binary vector with non-zero coordinates corresponding to the non-zero coordinates of μ, see page 95
$\mathrm{cl}(\xi)$	arc-consistency closure of a binary vector ξ, see Definition 6.11 on page 99
$\mathcal{J}$	the set of indexes of the Lagrange dual vector for the MAP-inference problem; $\lvert\mathcal{J}\rvert$ is equal to the number of coordinates in the dual vector, see §6.1 on page 85
$\left\langle \frac{1}{2} \right\rangle$, $\langle 0.7 \rangle$	angular brackets are used in figures for coordinates of primal relaxed solutions, see e.g. Figure 4.1, 4.2, 6.4, 12.3

Full text available at: http://dx.doi.org/10.1561/0600000084

1

Introduction to Inference for Graphical Models

There are many problems in computer science, which can be formulated in the form of so-called *Discrete Graphical Models*. Examples can be found in bio-informatics, communication theory, statistical physics, computer vision, signal processing, information retrieval and machine learning.

Discrete graphical models as a modeling tool naturally appear when

- the target object (the object we model) consists of *many small parts*,
- each part must be labeled by a label from a *finite set*, and
- parts (and, therefore, their labels) are *mutually dependent*.

Example 1.1 (Image segmentation). Image segmentation is a classical image analysis task: Each pixel of an input image must be assigned a label of an object visible in the image. For instance, if we consider images of street scenes, these labels could belong to the set {`pedestrian`, `car`, `tree`, `building`}.

The target object is an image, i.e. a two-dimensional array of pixels. Each pixel constitutes an elementary part of the image and must be

Full text available at: http://dx.doi.org/10.1561/0600000084

labeled with a label from a finite set. The simplest assumption about image segments, i.e. groups of pixels having the same label, is the so called "compactness assumption". It states that it is more probable that neighboring pixels are labeled with the same label than with different ones.

Example 1.2 (Depth reconstruction)**.** Depth reconstruction is another important image analysis problem. In the classical setting there are (at least) two images taken from different viewpoints. The task is to match pixels from these two images to each other. Assuming the positions of cameras and their focal lengths are known, this allows us to estimate depth of the scene, which was photographed with the cameras.

As in the previous example, the target object is a two-dimensional pixel array, where each pixel constitutes an elementary part of the object and must be labeled with a label from a finite set. Here, the meaning of the labels is different: Each label represents depth information of the associated pixel in an image, i.e. how far the depicted observation is placed from the camera. Usually the set of labels is chosen as natural numbers in a given interval, for instance, $\{0, 1, \dots, 255\}$.

Assuming that the observed surface is smooth, one would expect the difference $|s - s'|$ between labels s and s' in neighboring pixels to be small. The opposite would mean a jump in depth, or, in other words, non-smoothness of the surface.

Example 1.3 (Cell tracking problem in bio-imaging)**.** Given is a sequence of images that show the development of a living organism from an early embryo consisting of only a few cells to a fully grown animal. During this sequence, the images show moving and splitting cells.

Under the assumption that the image is already *pre-segmented*, i.e. the cells were already found in each image, the task at hand is to track each individual cell and its descendants from the first to the last frame.

The cells are the elementary parts of the considered object. Each cell in a given image frame is labeled with pointers to one or two cells in the next frame. One pointer means that the cell only moved, and two pointers correspond to a cell division. The simplest tracking model forbids two different cells to have the same descendants. This rule defines dependencies between object parts.

Full text available at: http://dx.doi.org/10.1561/0600000084

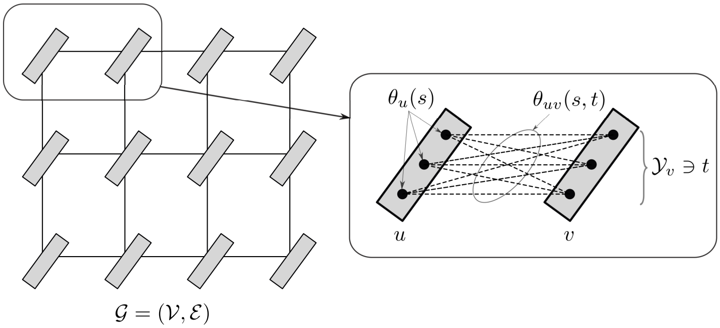

Figure 1.1: Example of a graphical model with grid structure. On the left, graph nodes are denoted with inclined rectangles, lines connecting nodes correspond to graph edges. On the right, two neighboring nodes are shown. Black circles inside rectangles correspond to the labels s in the node u (left rectangle) and t in the node v (right rectangle). Dashed lines correspond to each label pair s, t with an assigned pairwise cost $\theta_{uv}(s,t)$.

1.1 Basic definitions

Graph Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph consisting of a finite set of *nodes* $\mathcal{V}$ and a set of edges $\mathcal{E} \subseteq \binom{\mathcal{V}}{2}$. The set $\mathcal{E}$ will also be called *a neighborhood structure* of $\mathcal{V}$. For convenience, we will typically use lower case letters u and v for nodes of the graph, and write uv to denote an edge $\{u, v\} \in \mathcal{E}$ connecting u and v. Since the graph is undirected, uv and vu denote the same edge. The notation $\mathcal{N}_b(u)$ will be used for the set of nodes $\{v \mid uv \in \mathcal{E}\}$ connected to the node u.

The graph $\mathcal{G}$ is considered as a model of the considered target object, where the nodes represent the elementary object parts and edges stand for mutually dependencies between them.

In Examples 1.1 and 1.2 the graph $\mathcal{G}$ may have the grid structure of the underlying two-dimensional pixel array. In Example 1.3 cells of one image frame are neighbors, since their labels depend on each other.

Labels and unary costs A finite *set of labels* $\mathcal{Y}_u$ is associated with each node $u \in \mathcal{V}$. Our preference for each label is expressed by the *unary cost function* $\theta_u \colon \mathcal{Y}_u \to \mathbb{R}$, which is defined for each node $u \in \mathcal{V}$. The value $\theta_u(s)$ determines the *cost*, which we pay for assigning label $s \in \mathcal{Y}_u$ to the node u. Sometimes we will use very high costs to implicitly forbid

Full text available at: http://dx.doi.org/10.1561/0600000084

certain labels oder label pairs. The notation ∞ will be used to denote such high costs.

Unary costs are usually defined by what is known from observation. In Example 1.1, typically, the color distribution in the vicinity of a given pixel defines the cost of each possible label. The difference between color distributions from two or more images of the same scene taken from different viewpoints determines the unary costs in Example 1.2. Unary costs are often called the "data term" to emphasize that they depend on the input data or observation.

Dependence and pairwise costs Dependencies between labels assigned to different graph nodes are modeled with *pairwise cost functions* $\theta_{uv}\colon \mathcal{Y}_u \times \mathcal{Y}_v \to \mathbb{R}$, which are defined for each edge $uv \in \mathcal{E}$ of the graph.

A simple (although not always the best) way to model the compactness assumption in Example 1.1 is to assign

$$\theta_{uv}(s,t) = \begin{cases} 0, & s = t \\ \alpha, & s \neq t \end{cases} \tag{1.1}$$

for any pair of labels $(s,t) \in \mathcal{Y}_u \times \mathcal{Y}_v$ with some $\alpha > 0$. A simple way to model a smooth surface in depth reconstruction in Example 1.2 is to assign

$$\theta_{uv}(s,t) = |s - t|\,, \tag{1.2}$$

to penalize large differences between depth in the neighboring nodes.

In the cell tracking example the pairwise costs should forbid the same labels to be assigned to neighboring nodes when no cell division happens:

$$\theta_{uv}(s,t) = \begin{cases} 0, & s \neq t \\ \infty, & s = t\,. \end{cases} \tag{1.3}$$

This disallows that cells u and v "glue" to the same "parent" cell $s = t$. In case of cell division, this pairwise cost function can be extended in a natural way to disallow intersection of cell descendants.

These examples show that pairwise costs often incorporate the prior information about a considered object, therefore, they are often

Full text available at: http://dx.doi.org/10.1561/0600000084

collectively referred to as the *prior*. However, this is not always the case. For instance, much better segmentation results can be obtained if the parameter α in (1.1) depends on the color distribution of the input image, i.e. on uv.

Costs and cost functions are also called *potentials* and *potential functions*. We prefer the term *cost* since it is more widely used in general optimization literature.

Since unary and pairwise costs are functions of discrete variables, they can be seen as vectors. Therefore we can treat the unary cost function θ_u as a *unary cost* vector $(\theta_u(s),\ s \in \mathcal{Y}_u)$. Similar reasoning holds also for each pairwise cost function, which can be considered as a *pairwise cost* vector $\theta_{uv} = (\theta_{uv}(s,t),\ (s,t) \in \mathcal{Y}_u \times \mathcal{Y}_v)$. Unless we use the word *vector* or *function*, the context will determine whether we refer to a vector or a function θ_u (or θ_{uv}). All unary vectors stacked together form the vector of all unary costs $\theta_{\mathcal{V}} = (\theta_u,\ u \in \mathcal{V})$. The vector $\theta_{\mathcal{E}}$ of all pairwise costs is defined similarly as $(\theta_{uv},\ uv \in \mathcal{E})$. Stacking together the latter two results in a long *cost vector* $\theta = (\theta_{\mathcal{V}}, \theta_{\mathcal{E}})$ with dimension $\mathcal{I} := \sum_{u\in\mathcal{V}} |\mathcal{Y}_u| + \sum_{uv\in\mathcal{E}} |\mathcal{Y}_{uv}|$.

Labeling In the following, we will often use the notation $\mathcal{Y}_{\mathcal{A}}$ for all possible label assignments to a subset of nodes $\mathcal{A} \subseteq \mathcal{V}$. Formally, $\mathcal{Y}_{\mathcal{A}}$ stands for the Cartesian product $\prod_{u\in\mathcal{A}} \mathcal{Y}_u$. In particular, $\mathcal{Y}_{uv}$ denotes $\mathcal{Y}_u \times \mathcal{Y}_v$ and is the set of all possible pairs of labels in nodes u and v. A vector $y \in \mathcal{Y}_{\mathcal{V}}$ of labels assigned to *all* nodes of the graph is called *labeling*. We will refer to coordinates of this vector with the node index, i.e. y_u stands for the label assigned to the node u. One may also speak about *partial labelings*, if only a subset $\mathcal{A}$ of the nodes is labeled.

Definition 1.4 (Graphical model)**.** The triple $(\mathcal{G}, \mathcal{Y}_{\mathcal{V}}, \theta)$ consisting of a graph $\mathcal{G}$, discrete space of all labelings $\mathcal{Y}_{\mathcal{V}}$ and a corresponding cost vector θ, is called *a graphical model.*

Definition 1.5 (Energy minimization problem)**.** The problem

$$y^* = \arg\min_{y\in\mathcal{Y}_{\mathcal{V}}} \left[E(y;\theta) := \sum_{u\in\mathcal{V}} \theta_u(y_u) + \sum_{uv\in\mathcal{E}} \theta_{uv}(y_u, y_v) \right] \tag{1.4}$$

Full text available at: http://dx.doi.org/10.1561/0600000084

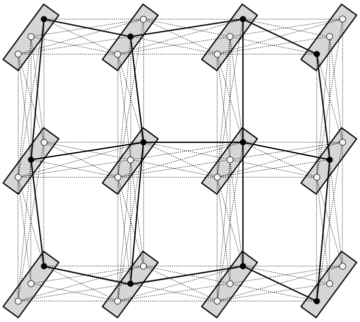

Figure 1.2: Labeling of the graphical model from Figure 1.1. Selected labels are marked as black circles and connected with solid lines. Each black circle corresponds to a unary cost and each solid line to a pairwise cost in the sum in the energy minimization problem (1.4).

of finding a labeling y^* with minimal total cost will be called *energy minimization* or *maximum a posteriori (MAP) inference* problem for the graphical model $(\mathcal{G}, \mathcal{Y}_\mathcal{V}, \theta)$.

For the sake of notation we will sometimes use the short form of (1.4)

$$y^* = \arg\min_{y \in \mathcal{Y}_\mathcal{V}} \left[E(y;\theta) := \sum_{w \in \mathcal{V} \cup \mathcal{E}} \theta_w(y_w) \right] \tag{1.5}$$

with y_w being equal to y_u, if w corresponds to a node, i.e. $w = u \in \mathcal{V}$, and y_{uv}, if w corresponds to an edge, i.e. $w = uv \in \mathcal{E}$.

Problems equivalent or very closely related to (1.4) have also other names depending on the corresponding community they are studied in: *maximum likelihood explanation (MLE) inference* (machine learning, natural language processing community), *weighted/valued/partial constraint satisfaction problem* (constraint satisfaction community).

1.2 Probabilistic interpretation

The name *MAP-inference* stems from the probabilistic interpretation of the problem (1.4). With the energy $E(y;\theta)$ one typically associates the exponential probability distribution

$$p(y) = \frac{1}{Z(\theta)} \exp\left(-E(y;\theta)\right), \tag{1.6}$$

Full text available at: http://dx.doi.org/10.1561/0600000084

where the normalizer $Z(\theta)$ is known as *partition function*.

According to the distribution (1.6), problem (1.4) is equivalent to finding *the most probable* labeling y, i.e. the one maximizing $p(y)$. Since E has the separable form (1.5) the expression (1.6) takes the form of the *Gibbs distribution*

$$p(y) = \frac{1}{Z(\theta)} \prod_{w \in \mathcal{V} \cup \mathcal{E}} \Theta_w(y_w) \tag{1.7}$$

with $\Theta_w = \exp(-\theta_w)$. This explains the also frequently used name "factors" for the cost functions and their exponents Θ_w.

The probabilistic interpretation (1.6) gives rise to several other *probabilistic inference* problems motivated by Bayesian statistical decision making theory. One computational problem, often referred to as *marginalization inference*, consists of computing *marginal* distributions

$$\hat{p}_u(s) := \sum_{y \in \mathcal{Y}_{\mathcal{V}} \colon y_u = s} p(y) \tag{1.8}$$

for each node u and label s of a graphical model. These kinds of problems, although closely related to MAP-inference, are beyond the scope of this monograph.

1.3 Combinatorial complexity of MAP-inference

The number of possible labelings y in (1.4) grows exponentially with the cardinality of $\mathcal{V}$, as it is equals $\prod_{v \in \mathcal{V}} |\mathcal{Y}_v|$. It results in $L^{|\mathcal{V}|}$ in case all nodes have the same number of labels $|\mathcal{Y}_u| = L$, $\forall u \in \mathcal{V}$.

However, an exponentially large set of solutions is not sufficient for polynomial $\mathcal{NP}$-hardness of a problem. For example, the shortest path between two nodes in a directed graph with positive edge weights has an exponentially large set of solutions, but is polynomially solvable by Dijkstra's algorithm.

Below we show that the MAP-inference (1.4) is, indeed, $\mathcal{NP}$-hard. To do so, it is sufficient to show that some $\mathcal{NP}$-complete decision problem is polynomially reducible to MAP-inference.

In the following construction we will show that the Hamiltonian cycle problem reduces to MAP-inference in polynomial time.

Full text available at: http://dx.doi.org/10.1561/0600000084

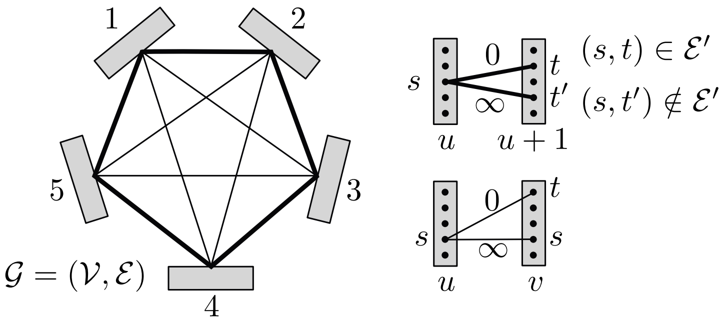

Figure 1.3: Illustration of the reduction of the Hamiltonian cycle problem to MAP-inference for a graph with 5 nodes. Edges of the MAP-inference graph $\mathcal{G}$ are divided into two groups: between nodes u and $u+1$ (bold edges) and all others (thin edges). The corresponding pairwise costs are illustrated on the right (see also main text).

Definition 1.6 (Hamiltonian cycle). A *Hamiltonian cycle* in a graph $\mathcal{G}$ is a cycle which visits each node exactly once.

The problem of deciding whether a given directed graph has a Hamiltonian cycle is known to be $\mathcal{NP}$-complete. To show $\mathcal{NP}$-hardness of MAP-inference, it is sufficient to reduce the Hamiltonian cycle problem to it.

Let $\mathcal{G}' = (\mathcal{V}', \mathcal{E}')$ be the graph for which one should solve the Hamiltonian cycle problem. Let us construct the following graphical model (see Figure 1.3): For the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ it holds that $\mathcal{V} = \mathcal{V}'$ and $\mathcal{E} = \binom{\mathcal{V}}{2}$. In other words, graph $\mathcal{G}$ contains the same nodes as graph $\mathcal{G}'$ and is fully connected. Moreover, we will order all nodes of the graph $\mathcal{G}$, i.e. $\mathcal{V} = \{1, 2, \dots, |\mathcal{V}|\}$. This order is the order of nodes in the Hamiltonian cycle we are searching for. We will assume the operation $u+1$ to be defined modulo $|\mathcal{V}|$, i.e. $u+1$ defines the next element of the Hamiltonian cycle. In other words, if $u < |\mathcal{V}|$ then $u+1$ is the next natural number after u and for $u = |\mathcal{V}|$ the element $u+1$ is equal to 1.

The set of labels $\mathcal{Y}_u := \mathcal{V}'$ is the same for each node $u \in \mathcal{V}$. Its elements index nodes of the graph $\mathcal{G}'$. A label s assigned to a node $u \in \mathcal{V}$ encodes that the u-th node in the Hamiltonian cycle corresponds to the node s of the graph $\mathcal{G}'$.

Full text available at: http://dx.doi.org/10.1561/0600000084

Unary costs are equal to 0. Pairwise costs are split into two groups. For a pair of nodes $\{u, u+1\} \in \mathcal{E}$ the cost reads

$$\theta_{u,u+1}(s,t) = \begin{cases} 0, & (s,t) \in \mathcal{E}' \\ \infty, & (s,t) \notin \mathcal{E}' \,. \end{cases} \tag{1.9}$$

It guarantees that two neighboring nodes of the Hamiltonian cycle are connected by an edge in the graph $\mathcal{G}'$.

To guarantee that no node is included twice in the Hamiltonian cycle, we set up other pairwise costs for $v \neq u+1$ and $u \neq v+1$ as follows:

$$\theta_{uv}(s,t) = \begin{cases} 0, & s \neq t \\ \infty, & s = t \end{cases} . \tag{1.10}$$

Such type of pairwise costs is sometimes called the *uniqueness constraints*, since these costs enforce that each label is selected at most ones.

Let y be some labeling of the graphical model $\mathcal{G}$ such that $E(y,\theta) < \infty$. Then the sequence $(y_1, y_2, \dots, y_{|\mathcal{V}|})$ is the Hamiltonian cycle by construction: there is an edge between y_u and y_{u+1} in $\mathcal{G}'$, and the set $\{y_1, y_2, \dots, y_{|\mathcal{V}|}\}$ is exactly the set $\mathcal{V}'$.

All labelings have either value 0 or ∞. Therefore, the solution of the MAP-inference problem answers the question whether there is a labeling y such that $E(y,\theta) < \infty$, and, therefore, whether there is a Hamiltonian cycle in the graph $\mathcal{G}'$.

Note that the same reduction of the Hamiltonian cycle problem to the MAP-inference could have also be done without using the infinite costs. Instead, any positive finite cost (e.g. 1) could be used in place of infinities. In this case the solution of the MAP-inference problem answers the question whether there is a labeling y such that $E(y,\theta) = 0$, which is equivalent to the existence of a Hamiltonian cycle in the graph $\mathcal{G}'$.

1.4 Bibliography and further reading

For further examples of applications of graphical models in computer vision and image processing we refer to the collection [10]. Books [77, 47] can be recommended to learn more about the probabilistic view on

Full text available at: http://dx.doi.org/10.1561/0600000084

graphical models. The monograph [135] concentrates on the exponential family (1.6) of distributions and its relation to graphical models.

A classical source to learn about the computational complexity of combinatorial problems is [27], a modern exposition is given in [5]. The most recent and comprehensive analysis of complexity of the MAP-inference problem is provided in [70].

The text books [25] and [103] can be recommended to learn about Bayesian decision theory.

The reduction of the Hamiltonian cycle problem to MAP-inference is reproduced from the lectures on structural pattern recognition given by Prof. Michail Schlesinger at National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" where the author studied mathematics and computer science.

Full text available at: http://dx.doi.org/10.1561/0600000084

References

[1] Aggarwal, A. and J. Park. 1989. "Sequential searching in multidimensional monotone arrays". *Research Report* No. RC 15128. Yorktown Heights, NY: IBM TJ Watson Research Center.

[2] Ahuja, R., D. Hochbaum, and J. Orlin. 2004. "A cut-based algorithm for the nonlinear dual of the minimum cost network flow problem". *Algorithmica*. 39(3): 189–208. UC Berkley manuscript (1999).

[3] Alahari, K., P. Kohli, and P. H. Torr. 2008. "Reduce, reuse & recycle: Efficiently solving multi-label MRFs". In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE. 1–8.

[4] Andres, B., J. H. Kappes, T. Beier, U. Köthe, and F. A. Hamprecht. 2012. "The lazy flipper: Efficient depth-limited exhaustive search in discrete graphical models". In: *European Conference on Computer Vision*. Springer. 154–166.

[5] Arora, S. and B. Barak. 2009. *Computational complexity: a modern approach*. Cambridge University Press.

[6] Beck, A. and L. Tetruashvili. 2013. "On the convergence of block coordinate descent type methods". *SIAM journal on Optimization*. 23(4): 2037–2060.

[7] Bellman, R. and S. Dreyfus. 1962. *Applied dynamic programming*. Princeton University Press.

Full text available at: http://dx.doi.org/10.1561/0600000084

[8] Bertsekas, D. P. 1999. *Nonlinear programming, second edition.* Athena scientific.

[9] Besag, J. 1974. "Spatial interaction and the statistical analysis of lattice systems". *Journal of the Royal Statistical Society. Series B (Methodological)*: 192–236.

[10] Blake, A., P. Kohli, and C. Rother. 2011. *Markov random fields for vision and image processing.* MIT Press.

[11] Bodlaender, H. L. and A. M. Koster. 2010. "Treewidth computations I. Upper bounds". *Information and Computation.* 208(3): 259–275.

[12] Bodlaender, H. L. and A. M. Koster. 2011. "Treewidth computations II. Lower bounds". *Information and Computation.* 209(7): 1103–1119.

[13] Boros, E. and P. L. Hammer. 2002. "Pseudo-boolean optimization". *Discrete applied mathematics.* 123(1): 155–225.

[14] Boros, E., P. Hammer, and X. Sun. 1991. "Network flows and minimization of quadratic pseudo-Boolean functions". *Tech. rep.*

[15] Boyd, S. and L. Vandenberghe. 2004. *Convex optimization.* Cambridge University Press.

[16] Boykov, Y., O. Veksler, and R. Zabih. 2001. "Fast approximate energy minimization via graph cuts". *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* 23(11): 1222–1239.

[17] Bulatov, A., P. Jeavons, and A. Krokhin. 2005. "Classifying the complexity of constraints using finite algebras". *SIAM Journal on Computing.* 34(3): 720–742.

[18] Burkard, R. E., B. Klinz, and R. Rudolf. 1996. "Perspectives of Monge properties in optimization". *Discrete Applied Mathematics.* 70(2): 95–161.

[19] Chen, Q. and V. Koltun. 2014. "Fast MRF optimization with application to depth reconstruction". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 3914–3921.

[20] Cohen, D. and P. Jeavons. 2006. "Chapter 8: The complexity of constraint languages". In: *Handbook of constraint programming.* Elsevier.

Full text available at: http://dx.doi.org/10.1561/0600000084

[21] Cooper, M. C., S. De Givry, M. Sánchez, T. Schiex, M. Zytnicki, and T. Werner. 2010. "Soft arc consistency revisited". *Artificial Intelligence*. 174(7-8): 449–478.

[22] Werner, T. 2009a. "Revisiting the Decomposition Approach to Inference in Exponential Families and Graphical Models". *Tech. rep.* CMP, Czech TU.

[23] Dantzig, G. B. and M. N. Thapa. 2006. *Linear programming 1: Introduction*. Springer Science & Business Media.

[24] Davey, B. A. and H. A. Priestley. 2002. *Introduction to lattices and order*. Cambridge University Press.

[25] Duda, R. O., P. E. Hart, and D. G. Stork. 2012. *Pattern classification*. John Wiley & Sons.

[26] Fu, Q. and H. W. A. Banerjee. 2013. "Bethe-ADMM for Tree Decomposition based Parallel MAP Inference". In: *UAI*.

[27] Garey, M. R. and D. S. Johnson. 2002. *Computers and intractability*. Vol. 29. WH Freeman New York.

[28] Globerson, A. and T. S. Jaakkola. 2008. "Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations". In: *Advances in Neural Information Processing Systems 20*.

[29] Greig, D. M., B. T. Porteous, and A. H. Seheult. 1989. "Exact maximum a posteriori estimation for binary images". *Journal of the Royal Statistical Society. Series B (Methodological)*: 271–279.

[30] Guignard, M. 2003. "Lagrangean relaxation". *Top*. 11(2): 151–200.

[31] Guignard, M. and S. Kim. 1987a. "Lagrangean decomposition for integer programming: theory and applications". *Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle*. 21(4): 307–323.

[32] Guignard, M. and S. Kim. 1987b. "Lagrangean decomposition: A model yielding stronger Lagrangean bounds". *Mathematical programming*. 39(2): 215–228.

[33] Haller, S., P. Swoboda, and B. Savchynskyy. 2018. "Exact MAP-Inference by Confining Combinatorial Search with LP Relaxation". In: *In Proceedings of AAAI*.

Full text available at: http://dx.doi.org/10.1561/0600000084

[34] Hammer, P. L., P. Hansen, and B. Simeone. 1984. "Roof duality, complementation and persistency in quadratic 0–1 optimization". *Mathematical programming*. 28(2): 121–155.

[35] Hazan, T. and A. Shashua. 2010. "Norm-product belief propagation: Primal-dual message-passing for approximate inference". *IEEE Transactions on Information Theory*. 56(12): 6294–6316.

[36] Ishikawa, H. 2003. "Exact optimization for Markov random fields with convex priors". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 25(10): 1333–1336.

[37] Johnson, J. K., D. M. Malioutov, and A. S. Willsky. 2007a. "Lagrangian Relaxation for MAP Estimation in Graphical Models". *CoRR*. abs/0710.0013. URL: http://arxiv.org/abs/0710.0013.

[38] Johnson, J. K., D. Malioutov, and A. S. Willsky. 2007b. "Lagrangian relaxation for MAP estimation in graphical models". In: *45th Ann. Allerton Conf. on Comm., Control and Comp.*

[39] Kainmueller, D., F. Jug, C. Rother, and G. Meyers. *Graph matching problems for annotating C. Elegans*. URL: https://datarep.app.ist.ac.at/57/.

[40] Kappes, J., B. Savchynskyy, and C. Schnörr. 2012. "A Bundle Approach To Efficient MAP-Inference by Lagrangian Relaxation". In: *CVPR 2012*.

[41] Kappes, J. H., B. Andres, F. A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, T. Kröger, J. Lellmann, N. Komodakis, B. Savchynskyy, and C. Rother. 2015. "A Comparative Study of Modern Inference Techniques for Structured Discrete Energy Minimization Problems". English. *International Journal of Computer Vision*: 1–30. DOI: 10.1007/s11263-015-0809-x. URL: http://dx.doi.org/10.1007/s11263-015-0809-x.

[42] Kappes, J. H., S. Schmidt, and C. Schnörr. 2010. "MRF Inference by k-Fan Decomposition and Tight Lagrangian Relaxation". In: *European Conference on Computer Vision (ECCV)*. Ed. by K. Daniilidis, P. Maragos, and N. Paragios. Vol. 6313. Berlin/Heidelberg: Springer. 735–747.

Full text available at: http://dx.doi.org/10.1561/0600000084

[43] Kappes, J. H., M. Speth, G. Reinelt, and C. Schnörr. 2013. "Towards efficient and exact MAP-inference for large scale discrete computer vision problems via combinatorial optimization". In: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE. 1752–1758.

[44] Kelm, B. M., N. Mueller, B. H. Menze, and F. A. Hamprecht. 2006. "Bayesian estimation of smooth parameter maps for dynamic contrast-enhanced MR images with block-ICM". In: *Computer Vision and Pattern Recognition Workshop, CVPRW'06. Conference on*. IEEE. 96–96.

[45] Kleinberg, J. and E. Tardos. 2002. "Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields". *Journal of the ACM (JACM)*. 49(5): 616–639.

[46] Kohli, P., A. Shekhovtsov, C. Rother, V. Kolmogorov, and P. Torr. 2008. "On partial optimality in multi-label MRFs". In: *Proceedings of the 25th international conference on Machine learning*. ACM. 480–487.

[47] Koller, D. and N. Friedman. 2009. *Probabilistic graphical models: principles and techniques*. MIT press.

[48] Kolmogorov, V. 2005. "Convergent tree-reweighted message passing for energy minimization". In: *AISTATS*.

[49] Kolmogorov, V. 2006. "Convergent tree-reweighted message passing for energy minimization". *IEEE transactions on pattern analysis and machine intelligence*. 28(10): 1568–1583.

[50] Kolmogorov, V. 2010. "Generalized roof duality and bisubmodular functions". In: *Advances in neural information processing systems*. 1144–1152.

[51] Kolmogorov, V. 2015. "A new look at reweighted message passing". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 37(5): 919–930.

[52] Kolmogorov, V. and C. Rother. 2007. "Minimizing nonsubmodular functions with graph cuts-a review". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 29(7): 1274–1279.

Full text available at: http://dx.doi.org/10.1561/0600000084

[53] Kolmogorov, V. and M. Wainwright. 2005. "On the optimality of tree-reweighted max-product message-passing". In: *Uncertainty in Artificial Intelligence (UAI)*.

[54] Kolmogorov, V. and R. Zabin. 2004. "What energy functions can be minimized via graph cuts?" *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 26(2): 147–159.

[55] Komodakis, N., N. Paragios, and G. Tziritas. 2007. "MRF optimization via dual decomposition: Message-passing revisited". In: *ICCV*.

[56] Komodakis, N., N. Paragios, and G. Tziritas. 2011. "MRF Energy Minimization and Beyond via Dual Decomposition". *IEEE Trans. PAMI*. 33(3): 531–552. DOI: http://dx.doi.org/10.1109/TPAMI.2010.108.

[57] Korte, B., J. Vygen, B. Korte, and J. Vygen. 2002. *Combinatorial optimization*. Springer.

[58] Koster, A. M., S. P. van Hoesel, and A. W. Kolen. 1998. "The partial constraint satisfaction problem: Facets and lifting theorems". *Operations Research Letters*. 23(3): 89–97. DOI: https://doi.org/10.1016/S0167-6377(98)00043-1. URL: http://www.sciencedirect.com/science/article/pii/S0167637798000431.

[59] Koval', V. K. and M. I. Schlesinger. 1976. "Dvumernoe programmirovanie v zadachakh analiza izobrazheniy (Two-dimensional programming in image analysis problems)". *Avtomatika i Telemekhanika*. (8): 149–168.

[60] Kovtun, I. 2003. "Partial Optimal Labeling Search for a NP-Hard Subclass of (max,+) Problems". In: *DAGM*.

[61] Kovtun, I. 2011. "Sufficient condition for partial optimality for (max, +) labeling problems and its usage". *Control Systems and Computers*. 2. Special issue.

[62] Kovtun, I. 2004. "Segmentaciya zobrazhen na usnovi dostatnikh umov optimalnosti v NP-povnikh klasakh zadach strukturnoi rozmitki (Image segmentation based on sufficient conditions of optimality in NP-complete classes of structural labeling problems)". *PhD thesis*. PhD thesis, IRTC ITS Nat. Academy of Science Ukraine, Kiev. In Ukrainian.

Full text available at: http://dx.doi.org/10.1561/0600000084

[63] Kschischang, F. R., B. J. Frey, and H.-A. Loeliger. 2001. “Factor graphs and the sum-product algorithm”. *IEEE Transactions on information theory*. 47(2): 498–519.

[64] Kumar, M. P., V. Kolmogorov, and P. H. Torr. 2009. “An analysis of convex relaxations for MAP estimation of discrete MRFs”. *Journal of machine learning research*. 10(Jan): 71–106.

[65] Kumar, M. P., O. Veksler, and P. H. Torr. 2011. “Improved moves for truncated convex models”. *Journal of machine learning research*. 12(Jan): 31–67.

[66] Lang, H., D. Sontag, and A. Vijayaraghavan. 2018. “Optimality of Approximate Inference Algorithms on Stable Instances”. In: *International Conference on Artificial Intelligence and Statistics*. 1157–1166.

[67] Lauritzen, S. L. 1996. *Graphical models*. Vol. 17. Clarendon Press.

[68] Lempitsky, V., S. Roth, and C. Rother. 2008. “Fusionflow: Discrete-continuous optimization for optical flow estimation”. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE. 1–8.

[69] Lempitsky, V., C. Rother, S. Roth, and A. Blake. 2010. “Fusion moves for Markov random field optimization”. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 32(8): 1392–1405.

[70] Li, M., A. Shekhovtsov, and D. Huber. 2016. “Complexity of discrete energy minimization problems”. In: *European Conference on Computer Vision*. Springer. 834–852.

[71] Luo, Z.-Q. and P. Tseng. 1992. “On the convergence of the coordinate descent method for convex differentiable minimization”. *Journal of Optimization Theory and Applications*. 72(1): 7–35.

[72] Luong, D., P. Parpas, D. Rueckert, and B. Rustem. 2012. “Solving MRF Minimization by Mirror Descent”. In: *Advances in Visual Computing*. Vol. 7431. Springer Berlin Heidelberg. 587–598.

[73] Martins, A. F. T., M. A. T. Figueiredo, P. M. Q. Aguiar, N. A. Smith, and E. P. Xing. 2011. “An Augmented Lagrangian Approach to Constrained MAP Inference”. In: *ICML*.

Full text available at: http://dx.doi.org/10.1561/0600000084

[74] McCormick, S. T. 2005. "Submodular Function Minimization". In: *Discrete Optimization*. Ed. by K. Aardal, G. Nemhauser, and W. R. Vol. 12. *Handbooks in Operations Research and Management Science*. Elsevier. 321–391. DOI: https://doi.org/10.1016/S0927-0507(05)12007-6. URL: http://www.sciencedirect.com/science/article/pii/S0927050705120076.

[75] Meshi, O. and A. Globerson. 2011. "An Alternating Direction Method for Dual MAP LP Relaxation". In: *ECML/PKDD (2)*. 470–483.

[76] Nesterov, Y. 2013. *Introductory lectures on convex optimization: A basic course*. Vol. 87. Springer Science & Business Media.

[77] Nowozin, S. and C. H. Lampert. 2011. "Structured learning and prediction in computer vision". *Foundations and Trends® in Computer Graphics and Vision*. 6(3–4): 185–365.

[78] Papadimitriou, C. H. and K. Steiglitz. 1982a. *Combinatorial optimization: Algorithms and complexity*. Courier Corporation.

[79] Papadimitriou, C. and K. Steiglitz. 1982b. *Combinatorial optimization: algorithms and complexity*. Prentice Hall Inc.

[80] Pearl, J. 1988. *Probabilistic reasoning in intelligent systems*. Palo Alto: Morgan Kaufmann.

[81] Polyak, B. T. 1987. *Introduction to optimization*. Optimization Software New York.

[82] Powell, M. J. 1973. "On search directions for minimization algorithms". *Mathematical Programming*. 4(1): 193–201.

[83] Prusa, D. and T. Werner. 2015. "Universality of the Local Marginal Polytope." *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 37(4): 898.

[84] Průša, D. and T. Werner. 2017a. "LP Relaxation of the Potts Labeling Problem Is as Hard as Any Linear Program". *IEEE Trans. Pattern Anal. Mach. Intell.* 39(7): 1469–1475.

[85] Průša, D. and T. Werner. 2013. "Universality of the Local Marginal Polytope". In: *Conf. on Computer Vision and Pattern Recognition*. IEEE Computer Society. 1738–1743.

Full text available at: http://dx.doi.org/10.1561/0600000084

[86] Průša, D. and T. Werner. 2015. "How Hard is the LP Relaxation of the Potts Min-Sum Labeling Problem?" In: *Conf. on Energy Minimization Methods in Computer Vision and Pattern Recognition, Hongkong*. Springer. 57–70.

[87] Průša, D. and T. Werner. 2017b. "LP Relaxations of Some NP-Hard Problems Are as Hard as any LP". In: *ACM-SIAM Symp. on Discrete Algorithm (SODA)*. SIAM. 1372–1382.

[88] Ravikumar, P., A. Agarwal, and M. Wainwright. 2010. "Message-passing for Graph-structured Linear Programs: Proximal Methods and Rounding Schemes". *JMLR*. 11: 1043–1080.

[89] Rockafellar, R. T. 1974. *Conjugate duality and optimization*. Vol. 16. SIAM.

[90] Rockafellar, R. T. 2015. *Convex analysis*. Princeton university press.

[91] Rosenfeld, A., R. A. Hummel, and S. W. Zucker. 1976. "Scene labeling by relaxation operations". *IEEE Transactions on Systems, Man and Cybernetics*. (6): 420–433.

[92] Rossi, F., P. Van Beek, and T. Walsh. 2006. *Handbook of constraint programming*. Elsevier.

[93] Rother, C., V. Kolmogorov, V. Lempitsky, and M. Szummer. 2007. "Optimizing binary MRFs via extended roof duality". In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE. 1–8.

[94] Savchynskyy, B., S. Schmidt, J. Kappes, and C. Schnörr. 2012. "Efficient MRF Energy Minimization via Adaptive Diminishing Smoothing". In: *UAI*. 746–755.

[95] Savchynskyy, B., J. H. Kappes, P. Swoboda, and C. Schnörr. 2013. "Global MAP-optimality by shrinking the combinatorial search area with convex relaxation". In: *Advances in Neural Information Processing Systems*. 1950–1958.

[96] Savchynskyy, B., J. Kappes, S. Schmidt, and C. Schnörr. 2011. "A Study of Nesterov's Scheme for Lagrangian Decomposition and MAP Labeling". In: *CVPR*.

[97] Savchynskyy, B. and S. Schmidt. 2014. "Getting Feasible Variable Estimates From Infeasible Ones: MRF Local Polytope Study". In: *Advanced Structured Prediction*. MIT Press.

Full text available at: http://dx.doi.org/10.1561/0600000084

[98] Schlesinger, D. 2007. "Exact solution of permuted submodular minsum problems". In: *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer. 28–38.

[99] Schlesinger, D. and B. Flach. 2006. "Transforming an arbitrary minsum problem into a binary one". *Tech. rep.*

[100] Schlesinger, M. I. 1999-2002. "Lectures on labeling problems attended by the author". Kyiv.

[101] Schlesinger, M. I. and V. V. Giginyak. 2007. "Solution to Structural Recognition (MAX,+)-problems by their Equivalent Transformations. in 2 Parts". *Control Systems and Computers*. (1-2).

[102] Schlesinger, M. and K. Antoniuk. 2011. "Diffusion algorithms and structural recognition optimization problems". *Cybernetics and Systems Analysis*. 47(2): 175–192.

[103] Schlesinger, M. and V. Hlavác. 2013. *Ten lectures on statistical and structural pattern recognition*. Vol. 24. Springer Science & Business Media.

[104] Schlesinger, M. and B. Flach. 2000. "Some solvable subclasses of structural recognition problems". In: *Czech Pattern Recognition Workshop*. Vol. 2000. 55–62.

[105] Schmidt, S., B. Savchynskyy, J. Kappes, and C. Schnörr. 2011. "Evaluation of a First-Order Primal-Dual Algorithm for MRF Energy Minimization". In: *EMMCVPR 2011*.

[106] Schoenemann, T., V. Kolmogorov, S. Nowozin, P. Gehler, J. Jancsary, and C. Lampert. 2014. "Generalized sequential tree-reweighted message passing". *Advanced Structured Prediction*: 75.

[107] Schrijver, A. 1998. *Theory of linear and integer programming*. John Wiley & Sons.

[108] Shekhovtsov, A. 2013. "Exact and Partial Energy Minimization in Computer Vision". *PhD Thesis CTU–CMP–2013–24*. CMP, Czech Technical University in Prague.

[109] Shekhovtsov, A., P. Swoboda, and B. Savchynskyy. 2017. "Maximum persistency via iterative relaxed inference with graphical models". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 40. No. 7. 1668–1682. DOI: 10.1109/TPAMI.2017.2730884.

Full text available at: http://dx.doi.org/10.1561/0600000084

[110] Shekhovtsov, A. 2014. "Maximum persistency in energy minimization". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1162–1169.

[111] Shekhovtsov, A. 2016. "Higher order maximum persistency and comparison theorems". *Computer Vision and Image Understanding*. 143: 54–79.

[112] Shekhovtsov, A., C. Reinbacher, G. Graber, and T. Pock. 2016. "Solving Dense Image Matching in Real-Time using Discrete-Continuous Optimization". In: *Proceedings of the 21st Computer Vision Winter Workshop (CVWW)*. 13. ISBN: 978-3-85125-388-7.

[113] Shekhovtsov, A., P. Swoboda, and B. Savchynskyy. 2015. "Maximum persistency via iterative relaxed inference with graphical models". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 521–529.

[114] Shlezinger, M. 1976. "Syntactic analysis of two-dimensional visual signals in the presence of noise". *Cybernetics and systems analysis*. 12(4): 612–628.

[115] Shor, N. Z. 2012. *Minimization methods for non-differentiable functions*. Vol. 3. Springer Science & Business Media.

[116] Sontag, D. A. 2010. "Approximate inference in graphical models using LP relaxations". *PhD thesis*. Massachusetts Institute of Technology.

[117] Sontag, D. and T. Jaakkola. 2009. "Tree block coordinate descent for MAP in graphical models". In: *Artificial Intelligence and Statistics*. 544–551.

[118] Sontag, D., T. Meltzer, A. Globerson, T. Jaakkola, and Y. Weiss. 2008. "Tightening LP relaxations for MAP using message passing". In: *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*. AUAI Press. 503–510.

[119] Storvik, G. and G. Dahl. 2000. "Lagrangian-based methods for finding MAP solutions for MRF models". *IEEE Transactions on Image Processing*. 9(3): 469–479.

Full text available at: http://dx.doi.org/10.1561/0600000084

[120] Swoboda, P., A. Shekhovtsov, J. Kappes, C. Schnörr, and B. Savchynskyy. 2016. "Partial optimality by pruning for MAP-inference with general graphical models". *IEEE Trans. Patt. Anal. Mach. Intell.* 38(7): 1370–1382. DOI: 10.1109/TPAMI.2015.2484327.

[121] Swoboda, P. and B. Andres. 2017. "A message passing algorithm for the minimum cost multicut problem". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 1617–1626.

[122] Swoboda, P., J. Kuske, and B. Savchynskyy. 2017a. "A dual ascent framework for Lagrangean decomposition of combinatorial problems". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 1596–1606.

[123] Swoboda, P., C. Rother, H. Abu Alhaija, D. Kainmuller, and B. Savchynskyy. 2017b. "A study of Lagrangean decompositions and dual ascent solvers for graph matching". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 1607–1616.

[124] Swoboda, P., B. Savchynskyy, J. Kappes, and C. Schnörr. 2014. "Partial optimality by pruning for MAP-inference with general graphical models". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 1170–1177.

[125] Swoboda, P., B. Savchynskyy, J. Kappes, and C. Schnörr. 2013. "Partial optimality via iterative pruning for the Potts model". In: *International Conference on Scale Space and Variational Methods in Computer Vision.* Springer. 477–488.

[126] Szeliski, R., R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. 2008. "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors". *Pattern Analysis and Machine Intelligence, IEEE Transactions on.* 30(6): 1068–1080.

[127] Thuerck, D., M. Waechter, S. Widmer, M. von Buelow, P. Seemann, M. Pfetsch, and M. Goesele. 2016. "A fast, massively parallel solver for large, irregular pairwise Markov random fields." In: *High Performance Graphics.* 173–183.

Full text available at: http://dx.doi.org/10.1561/0600000084

[128] Tourani, S., A. Shekhovtsov, C. Rother, and B. Savchynskyy. 2018. "MPLP++: Fast, Parallel Dual Block-Coordinate Ascent for Dense Graphical Models". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 251–267.

[129] Tseng, P. 1993. "Dual coordinate ascent methods for non-strictly convex minimization". *Mathematical programming*. 59(1-3): 231–247.

[130] Tseng, P. 2001. "Convergence of a block coordinate descent method for nondifferentiable minimization". *Journal of optimization theory and applications*. 109(3): 475–494.

[131] Tseng, P. and S. Yun. 2009. "Block-coordinate gradient descent method for linearly constrained nonsmooth separable optimization". *Journal of optimization theory and applications*. 140(3): 513.

[132] Viterbi, A. 1971. "Convolutional Codes and Their Performance in Communication Systems". *IEEE Transactions on Communication Technology*. 19(5): 751–772. DOI: 10.1109/TCOM.1971.1090700.

[133] Vodolazskii, E. V., B. Flach, and M. I. Schlesinger. 2014. "Minimax problems of discrete optimization invariant under majority operators". *Computational Mathematics and Mathematical Physics*. 54(8): 1327–1336.

[134] Wainwright, M. J., T. S. Jaakkola, and A. S. Willsky. 2005. "MAP estimation via agreement on trees: message-passing and linear programming". *IEEE transactions on Information Theory*. 51(11): 3697–3717.

[135] Wainwright, M. J. and M. I. Jordan. 2008. "Graphical models, exponential families, and variational inference". *Foundations and Trends® in Machine Learning*. 1(1-2): 1–305.

[136] Waltz, D. L. 1972. "Generating semantic description from drawings of scenes with shadows". *Tech. rep.* No. AI271. MIT Artificial Intelligence Laboratory.

[137] Wang, H. and D. Koller. 2013. "Subproblem-Tree Calibration: A Unified Approach to Max-Product Message Passing." In: *ICML (2)*. 190–198.

Full text available at: http://dx.doi.org/10.1561/0600000084

[138] Werner, T. 2009b. "Revisiting the linear programming relaxation approach to Gibbs energy minimization and weighted constraint satisfaction". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 32(8): 1474–1488.

[139] Werner, T. 2017. "On Coordinate Minimization of Convex Piecewise-Affine Functions". *arXiv preprint arXiv:1709.04989*.

[140] Werner, T. 2007. "A linear programming approach to max-sum problem: A review". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 29(7): 1165–1179.

[141] Werner, T. and D. Pruša. 2019. "Relative Interior Rule in Block-Coordinate Minimization". *arXiv preprint arXiv:1910.09488*.

[142] Zach, C. 2015. "A Novel Tree Block-Coordinate Method for MAP Inference". In: *German Conference on Pattern Recognition*. Springer. 320–330.