source: Shekhovtsov, A. “Supermodular decomposition of structural labeling problem”, Control Systems and Machienes, Kiev 2005, accepted, in Russian. http://cmp.felk.cvut.cz/~shekhovt/publications/shekhovtsov-SM-05-art-en.pdf

SUPERMODULAR DECOMPOSITION OF STRUCTURAL LABELING PROBLEM[1]

Alexander Shekhovtsov

International Research and Training Center for Information Technologies and Systems, Kiev Glushkova str. 40, UKR

Abstract. We consider the problem of selecting labels in vertices of the graph in order to maximize sum of weights, assigned to pairs of labels ((max,+) problem, MAP MRF, minimization of Gibbs energy). We construct upper bounds on the goal function as sum of two auxiliary super- and submodular problems. The tightest possible bound of this type can be found by solving a special case of LP. It's solution allows to fix some part of optimal configuration of the initial problem.

Introduction. In the pattern recognition we know fundamental problems, which arise as estimation of maximal a posteriori probability configuration of Markov Random Field, as search of the optimal labeling, as minimization of Gibbs energy. Optimization models of this type play an important role in other areas: in statistical mechanics they arise as minimization of Hamiltonian energy of spin glass systems, in communication theory as problems of optimal error-correcting decoding. Their abstract formulations are studied in operations research, discrete mathematics, combinatorial theory.

Problem formulation. Let $G=(T,E)$ be a directed graph, where $E \subseteq \binom{T}{2}$. Elements of set T we'll refer as objects, and elements of E as pairs. United set $\Im = T \cup E$ we'll call as *structure* of the problem. Let K be finite set of labels. Let us call mapping $k: T \to K$ a *labeling* ($k \in K^T$). By $k_t : t \to K$ ($k_t \in K^t$) we'll denote restriction of labeling k to the subset $t \in \Im$. Let $q_{t;k_t} \in \mathbb{R}$ be

[1] This paper was supported by the INTAS project PRINCESS 04-77-7347.

parameters, associated to each pair (t, k_t), where $t \in \Im$, and $k_t \in K^t$. Pairs of type (e, k_e), $e \in E$ we will refer as edges, and pairs (t, k_t), $t \in T$ as nodes.

We consider the problem of finding a labeling with best total quality of edges and nodes:

$$\operatorname*{argmax}_{k \in K^T} \sum_{t \in \Im} q_{t;k_t}. \tag{1}$$

Following the terminology of [1], where a general formulation of the problem in arbitrary semiring is considered, we will use notation of semiring operations to denote problem (1), namely $(max,+)$ problem. It is known that $(max,+)$ problem is an NP-complete one, and that it can be reformulated as searching the maximal cut in network (MAX-CUT). There exists "supermodular" subclass of $(max,+)$ problems, which is equivalent to (solvable) MAX-CUT problems with non positive edge capacities. This solvable subclass is widely exploited in modeling of some applied problems in pattern recognition and image analysis. Except of supermodular $(max,+)$, problems with tree structure of the graph are solvable by means of dynamic programming. The latter subclass is else widely applicable mostly because it's simplicity. Unfortunately, there is a lot of applied tasks, which we could model with $(\max,+)$ optimization, but they do not fit in before mentioned solvable subclasses.

In [2] the notion of equivalent $(\max,+)$ problems was proposed: problems, in which the same quality function is represented with different vectors of parameters q. Indeed, we could e.g. omit using node weights (set them to 0), and define the same quality function using edge weights only. It is shown, that if there exists such equivalent representation, where an optimal labeling can be constructed from locally maximal edges, i.e. edges $(e, k_e) = (e, \arg\max_{k_e} q_{e;k_e})$, $e \in E$, by polynomial algorithm, then the representation itself can be found in polynomial time. It is known [3] that subclass of problems solvable in this way, includes all the supermodular problems, as well as all tree-structured problems. Special algorithms for trivializing the $(max,+)$ problem were proposed in [4, 5], algorithms in [6, 7] can be else viewed as finding the trivial equivalent problem.

Recently techniques of constructing special auxiliary problems were proposed [8, 9], which allows to find a part of an optimal solution for general class of (max,+) problems. There exists methods for upper bounding the goal function [10, 11], which else allows to find a part of optimal solution for these hard problems. We introduce some new properties of supermodular $(max,+)$ problems, which allows us to construct another upper bound on the goal function, and to analyze the set of optimal solutions of the $(max,+)$ problem. Proposed technique, when applied to $(max,+)$ problems with $K = 2$, provides the same bound and part of optimal solution as [10]. There are some useful consequences for problems with $K > 2$. In particular our technique allows recovering partial optimality in the form of restricted intervals of labels in each object. For supermodular problems our approach gives an exact solution, this result is else in accordance with [10] and [3]. Unfortunately for chain-structured problems with $K > 2$ our approach does not provide an exact solution anymore. In more details these results of this paper outlined in [12].

Upper bound. Let us write problem (1) in the form of scalar product:

$$\max_{k\in K^T} \sum_{t\in\Im} q_{t;k_t} = \max_{k\in K^T} \sum_{t\in\Im} \sum_{k'_t} q_{t;k'_t} \mathbf{1}_{\{k_t = k'_t\}} = \max_{k\in K^T} \langle q, f(k) \rangle, \tag{2}$$

where $f(k)_{t;k'_t} = \mathbf{1}_{\{k_t = k'_t\}}$. Using this representation it is easy to see that the following inequality holds:

$$\max_{k\in K^T} \langle q, f(k) \rangle \le \sum_i \max_{k\in K^T} \langle q^i, f(k) \rangle, \text{ whenever } q = \sum_i q^i. \tag{3}$$

It establishes an upper bound on the quality of best labeling $\Phi(q) := \max_{k\in K^T} \langle q, f(k) \rangle$ as the sum of qualities of best labelings of auxiliary problems $\Phi(q^i) := \max_{k\in K^T} \langle q^i, f(k) \rangle$. The following lemma states a connection between solutions of auxiliary problems and solution of original problem, in the case when (3) holds as equality.

Lemm 1 ([6]). Inequality *(3) holds as equality, if and only if* $\bigcap_i OPT(q^i) \neq \varnothing$, *where*

$OPT(q^i) = \operatorname{Arg}\max_{k\in K^T} \langle q^i, f(k) \rangle$ *– set of optimal labelings for* q^i.

If the upper bound (3) provides q^i, satisfying $\exists k^* \in \bigcap_i OPT\,|(q^i)$, then k^* is an optimal solution of the original problem. Generally, evaluating the value $\bigcap_i OPT\,|\,(q^i) \neq \varnothing$ appears to be an NP-hard problem, however in some special cases it is known to be polynomially solvable.

Tree decomposition. Let us consider a decomposition $\sum_i q^i = q$, proposed in [6], when problems $\Phi(q^i) = \max_{k \in K^T} \langle q, f(k) \rangle$ constructed to appear solvable as tree-structured problems. Let for each i: (V_i, E_i) be a tree, which is a subgraph of G. Let $\vec{\Theta}^i$ be a set of parameter vectors, *respecting the structure of the tree*: $\vec{\Theta}^i = \{q^i \,|\, q^i_{t;k_t} = 0 \text{ for all } t \notin V_i \cup E_i, \text{and all } k_t\}$.

Each problem $\Phi(q^i)$ appears then to be easily solvable as $(max,+)$ problem on the tree (V_i, E_i).

A tree decomposition is searched, which provides the tightest possible bound:

$$\min_{\substack{q^i \in \vec{\Theta}^i \\ \sum_i q^i = q}} \sum_i \Phi(q^i) \geq \Phi(q). \tag{4}$$

Equality $\sum_i q^i = q$ must hold, therefore each nonzero element $q_{t;k_t}$ must be represented in at least one nonzero element $q^i_{t;k_t}$, from which it follows that trees (V_i, E_i) must cover the graph G. Following theorem gives a complete description of upper bound (4), which is independent of selection of trees (V_i, E_i) covering G.

Theorem 1 (LOCAL2 relaxation , [6]). *Lagrangian dual to minimization problem (4) is the maximization problem:* $\max_{m \in LOCAL2(G)} \langle q, m \rangle$, (5)

where $\mathrm{LOCAL\,2}(G)$ *— is a set of node-agreed distributions over* K^t :

$$\mathrm{LOCAL\,2}(G) = \left\{ m \,\middle|\, m_{t;k_t} \geq 0, \sum_{k_t} m_{t;k_t} = 1, \sum_{k_{t'}} m_{tt';k_{tt'}} = m_{t;k_t} \;\; \forall tt' \in E \right\}. \tag{6}$$

Set $\mathrm{LOCAL\,2}(G)$ is a convex polyhedron, restricted with polynomial number of linear constraints, namely $O(|\Im||K|^2)$ constraints, thus problem (5) is polynomially solvable as linear programming (LP) task, and consequently (4) is. Efficiency of general LP solvers is not sufficient for large scale problems. There were proposed several algorithms aimed at direct solution of (4). As far as it does not influence the value of (4) we can select the set of trees, covering the graph, as minimal spanning trees [6], or as chains [7], or as separate edges of G. In the last case it appears, that problem (5) can be else solved by means of equivalent transformations technique [2](See else further research in [3, 1]), proposed by Schlesinger in 1976, for which convergent algorithms were developed [4, 5] with stationary point properties fully analogues to that of [7].

Following representation of initial problem (2) in the form of linear programming task, shows that problem (5) is indeed a relaxation of (2).

Let $\mathcal{P}=\left\{p:K^T\to\mathbb{R}\,\middle|\,\sum_{k\in K^T}p(k)=1\,,p(k)\geq 0\right\}$ denote the set of distributions over K^T. The initial $(max,+)$ problem can be written in the form:

$$\max_{k\in K^T}\langle q,f(k)\rangle=\max_{p\in\mathcal{P}}\sum_k p(k)\langle q,f(k)\rangle, \tag{7}$$

where equality holds, because $\sum_k p(k)\langle q,f(k)\rangle\leq\max_{k\in K^T}\langle q,f(k)\rangle$ as convex combination of $\langle q,f(k)\rangle$, and, because $\sum_k p(k)\langle q,f(k)\rangle=\max_{k\in K^T}\langle q,f(k)\rangle$ for $p(k)=\mathbf{1}_{\{k=k^*\}}\in\mathcal{P}$.

Further, using the linearity of $\langle q,\cdot\rangle$:

$$\max_{p\in\mathcal{P}}\sum_k p(k)\langle q,f(k)\rangle=\max_{p\in\mathcal{P}}\left\langle q,\sum_k p(k)f(k)\right\rangle=\max_{m\in\mathrm{MARG}(G)}\langle q,m\rangle, \tag{8}$$

where $\mathrm{MARG}(G)=\left\{m\,\middle|\,\exists p\in\mathcal{P}:m=\sum_k p(k)f(k)\right\}$.

Let us note that for vector m from $\mathrm{MARG}(G)$ it holds: $m_{t;k'_t}=\sum_k p(k)f(k)_{t;k'_t}=\sum_k p(k)\mathbf{1}_{\{k_t=k'_t\}}$. Thus, $m_{t;(\cdot)}$ are defined as *marginal probabilities* of some distribution $p\in\mathcal{P}$, so we will refer vector m from $\mathrm{MARG}(G)$ as *marginal vector*. Set $\mathrm{MARG}(G)$ is a convex polyhedron, bounded with exponentially large number of linear constraints. It's easy to see that $\mathrm{MARG}(G)\subseteq\mathrm{LOCAL}\,2(G)$, i.e. some constraints defining $\mathrm{MARG}(G)$, are relaxed in $\mathrm{LOCAL}\,2(G)$. Therefore next inequality holds:

$$\max_{m\in \mathrm{MARG}(G)} \langle q, m\rangle \le \max_{m\in \mathrm{LOCAL2}(G)} \langle q, m\rangle . \tag{9}$$

Inequality (9) can be obtained else another way around, using duality:

$$\max_{m\in \mathrm{MARG}(G)} \langle q, m\rangle = \Phi(q) \le \min_{\substack{q^i\in\vec{\Theta}^i \\ \sum_i q^i = q}} \sum_i \Phi(q^i) = \max_{m\in \mathrm{LOCAL2}(G)} \langle q, m\rangle . \tag{10}$$

Supermodular (max,+) problems. Let U be a finite set. Function of subsets $F: 2^U \to \mathbb{R}$ is called *supermodular*, if: $\forall A, B \subseteq U:\ F(A) + F(B) \le F(A\cap B) + F(A\cup B)$. Definition of *submodular* function is different in inequality sign $\ge$.

It is known, that the problem of supermodular function maximization i.e. the problem $\operatorname{argmax}_{X\subseteq U} F(X)$, it polynomially solvable. On this basis we distinguish *supermodular* $(max,+)$ problems as such, that could be converted (in polynomial time) to problem of maximizing a supermodular function of subsets. Supermodular (max,+) problems of second order can be converted to MAX-CUT problem with nonpositive capacities[2] [13, 14, 15, 16, 17], for which there exists efficient algorithms [18, 19, 20]. On the other hand, any MAX-CUT problem with nonnegative capacities can be converted to supermodular (max,+) problem (e.g. [17]).

We will give the direct definition of supermodular $(max,+)$ problem, after [8].

Let set K be completely ordered. Let us define $k_t \sqcup k'_t := \max(k_t, k'_t)$, and $k_t \sqcap k'_t := \min(k_t, k'_t)$, where $t\in T$. Let set of labelings K^T be partially ordered with respect to order in K: for any two labels $k, k' \in K^T$ we will define maximal labeling $(k\sqcup k')_t := k_t \sqcup k'_t$, and minimal $(k\sqcap k')_t := k_t \sqcap k'_t$.

Let $Q(k) = \langle q, f(k)\rangle$ denote the quality of labeling.

Definition 1. *Function* $Q: K^T \to \mathbb{R}$ *is called supermodular, if:*

$\forall k, k' \in K^T:\quad Q(k) + Q(k') \le Q(k\sqcup k') + Q(k\sqcap k')$.

Theorem 2 ([21, 8, 9])**.** *Function* $Q: K^T \to \mathbb{R}$ *is supermodular if and only if:*

---

[2]The same as MIN-CUT problem with nonnegative edge capacities.

$$\forall \{t,t'\} \in E, \forall k_t <|K|, \forall k_{t'} <|K|: \quad \frac{\partial^2}{\partial k_t \partial k_{t'}} Q(k) \geq 0, \tag{11}$$

where $\frac{\partial}{\partial k_t} Q(k_{t_1}, \ldots k_t, \ldots k_{t_{|T|}}) = Q(k_{t_1}, \ldots k_t + 1, \ldots k_{t_{|T|}}) - Q(k_{t_1}, \ldots k_t, \ldots k_{t_{|T|}})$.

(In [21] this criterion is proposed for set functions $Q : 2^T \to \mathbb{R}$).

Corollary 1 ([8, 9])**.** *Let* $q = g + q$ *, where*[3] $g = q\,|_E$ *?* $q = q\,|_T$.

Function Q is supermodular $\Leftrightarrow$ $\forall \{t,t'\} \in E, \forall k_t <|K|, \forall k_{t'} <|K|: \quad \frac{\partial^2}{\partial k_t \partial k_{t'}} g_{\{t,t'\};k_{\{t,t'\}}} \geq 0.$

Supermodularity of Q is strongly dependent on the order selected in K. Instead of single set K, we will consider different sets K_t, associated witch each $t \in T$, and supplied their own order. Operations $\sqcap, \sqcup$ on labelings are defined as element-wise, so all the above stated remains intact.

Supermodular decomposition. Let $g : K^2 \to \mathbb{R}$ be function, defined on labels of single edge of graph G. Let *supM* denote the class of supermodular functions. Let *subM* denote the class of submodular functions.

Lemma 2. $\forall$ $g : K^2 \to \mathbb{R}$ $\exists g^1 \in supM$ *,* $\exists g^2 \in subM$ *:* $g = g^1 + g^2$.

Proof. From the representation $g_{ij} = \sum_{\substack{i'=i\ldots K-1 \\ j'=j\ldots K-1}} c_{i'j'} + a_i + b_j$, where

$$c_{ij} = \frac{\partial^2}{\partial i \partial j} g_{ij} = g_{ij} + g_{i+1\,j+1} - g_{i+1\,j} - g_{i\,j+1} \quad \forall i, j = 1 \ldots K-1,$$

$$a_i = g_{iK},\; b_j = g_{Kj} - g_{KK} \quad \forall i, j = 1 \ldots K,$$

we select $c_{ij}^1 = c_{ij} \mathbf{1}_{\{c_{ij} \geq 0\}}$ and $c_{ij}^2 = c_{ij} \mathbf{1}_{\{c_{ij} < 0\}}$, in order they satisfy $c_{ij}^1 + c_{ij}^2 = c_{ij}, \forall ij$.

Let us put[4]: $g_{ij}^1 = \sum_{\substack{i'=i\ldots K-1 \\ j'=j\ldots K-1}} c_{i'j'}^1 + a_i + b_j$, $\quad g_{ij}^2 = \sum_{\substack{i'=i\ldots K-1 \\ j'=j\ldots K-1}} c_{i'j'}^2$.

It can be directly verified that $g = g^1 + g^2$ and that $g^1 \in supM$ and $g^2 \in subM$. $\square$

[3] With $x\big|_A \in \mathbb{R}^3$ we will denote projection of x onto A, such that: $\forall t \in \Im :\; (x\big|_A)_t = x_t \mathbf{1}_{\{t \in T\}}$.

[4] How to share one-variable functions a and b between g^1 and g^2 does not matter.

Lemma 3 (Twisting)**.** $(g_{i,j}) \in subM \Rightarrow \left(g_{i,K-j}\right) \in supM$.

Proof. $g \in subM \quad \Rightarrow \quad \frac{\partial^2}{\partial i \partial j} g_{ij} \le 0,$ and so $\frac{\partial^2}{\partial i \partial j} g_{i,K-j} = -\frac{\partial^2}{\partial i \partial(-j)} g_{i,K-j} \ge 0.$ □

Corollary 2. If the graph G is twocolorable (or, what is the same, bipartite), then any submodular $(max,+)$ problem on G can be converted to supermodular, and thus can be solved in polynomial time. Indeed, for twocolorable graph we can reverse order in all odd vertices. All functions $g_e, e \in E,$ will convert from $subM$ to $supM$ by Lemma 3. This is an exclusive situation, because submodular (max,+) problems without restrictions on graph structure are NP-complete.

Upper bound with supermodular decomposition. Any decomposition q in a sum allows us to build an upper bound on the initial problem $\Phi(q)$. Let us consider the following, tightest among obtainable by Lemma 2, bound:

$$\min_{\substack{q^1+q^2=q \\ q^1 \in supM \\ q^2 \in subM}} \left[\Phi(q^1) + \Phi(q^2)\right] \ge \Phi(q). \tag{12}$$

Notice, that modification of parameters $q^i_{t;k}$, $t \in T,$ preserving the equality $q^1 + q^2 = q$ does not violate constraints in (12). Let us weaken (12) down to following:

$$\min_{\Delta q} \Phi(q^1 + \Delta q) + \Phi(q^2 - \Delta q) \ge \Phi(q), \tag{13}$$

where $\Delta q = \Delta q\big|_T$ – nodes values only, $q^1 \in supM, q^2 \in subM$, $q^1 + q^2 = q$.

Following, dual to (13), formulation will play an important role in consequent.

Theorem 3 (2MARG relaxation)**.** *Lagrangian dual to minimization problem (*13*) is the maximization problem:*

$$\max_{(m^1,m^2) \in \mathrm{2MARG}(G)} \left\langle q^1, m^1 \right\rangle + \left\langle q^2, m^2 \right\rangle, \tag{14}$$

$$\mathrm{2MARG}(G) = \left\{ (m^1, m^2) \,\middle|\, \begin{matrix} m^1 \in \mathrm{MARG}(G), \ m^2 \in \mathrm{MARG}(G) \\ m^1_{t;k'_t} = m^2_{t;k'_t} \quad \forall t \, \forall k'_t \end{matrix} \right\}. \tag{15}$$

Proof. Using strong duality theorem and some substitution of variables. □

Set of constraints 2MARG is of following form: m^1 and m^2 must be true marginal vectors and they must be node-agreed.

Minimum characterization. Let $q = q^1 + q^2$ be supermodular decomposition. Let $\mathrm{OPT}(q) = \mathrm{Arg}\max_{k \in K^T} \langle q, f(k) \rangle$. Let S be a set of labelings. Let $\mathcal{A}(S) = \{(t,l) \mid \exists k \in S : k_t = l\}$ denote the set of nodes, for which there exists a labeling from S passing through. Let $\Phi_{(t;l)}(q) = \max_{\substack{k:\\ k_t = l}} \langle q, f(k) \rangle$ denote max-marginal values for q in node (t,l). Let $\mathcal{A}(q) := \mathcal{A}(\mathrm{OPT}(q)) = \{(t,l) \mid \Phi_{(t;l)}(q) = \Phi(q)\}$ denote set of nodes, there exists an optimal labeling, passing through.

Statement 1 (Agreement on nodes)**.** *For the minimum of* $\Phi(q^1 - \Delta q) + \Phi(q^2 + \Delta q)$ *to be attained at* $\Delta q = 0$ *it is necessary that:*

$$\exists \mathrm{S1} \subseteq \mathrm{OPT}(q^1), \exists \mathrm{S2} \subseteq \mathrm{OPT}(q^2): \quad \mathcal{A}(\mathrm{S1}) = \mathcal{A}(\mathrm{S2}) \neq \varnothing. \tag{16}$$

Proof. Let $\mathcal{A}1 = \mathcal{A}(q^1)$ and let $\mathcal{A}2 = \mathcal{A}(q^2)$. We will modify (q^1, q^2) by means of the following procedure:

Suppose $\mathcal{A}1 \neq \mathcal{A}2$, consider, for certainty, that $\exists (t,l) \in \mathcal{A}1 : (t,l) \notin \mathcal{A}2$.

Then: $\Phi_{(t;l)}(q^1) = \Phi(q^1)$ and $\Phi_{(t;l)}(q^2) < \Phi(q^2)$.

There exists such $dq > 0$, that:

$$\begin{cases} \Phi_{(t;l)}(q^1 - dq|_{(t;l)}) = \Phi_{(t;l)}(q^1) - dq \\ \Phi_{(t;l)}(q^2 + dq|_{(t;l)}) < \Phi(q^2)\ . \end{cases} \tag{17}$$

Let us put $q^1{}_{(t;l)} -= dq$, $q^2{}_{(t;l)} += dq$. If $\Phi(q^1)$ has decreased, then we else decreased the sum $\Phi(q^1) + \Phi(q^2)$, which in contradiction to attainability of minimum at $\Delta q = 0$. Otherwise we get $\Phi_{(t;l)}(q^1) < \Phi(q^1)$ and thus set $\mathcal{A}1$ has shrunk at least by one node, when at the same time $\mathcal{A}2$ remained. Symmetrical considerations for the case of $\exists (t,l) \in \mathcal{A}2 : (t,l) \notin \mathcal{A}1$, will lead to that either $\Phi(q^2)$ will decrease, either $\mathcal{A}2$ will shrink.

Repeat (no more then $|T|(K-1)$ times) of this procedure will lead to uncovering a contradiction, or to stop with $\mathcal{A}1=\mathcal{A}2\neq\varnothing$ for some modified parameters $\tilde{q}^1,\tilde{q}^2$. Let then $\mathcal{A}:=\mathcal{A}(\tilde{q}^1)=\mathcal{A}(\tilde{q}^2)$. Let us put $S1=\mathrm{O}PT(\tilde{q}^1)$. By the construction of set $\mathcal{A}$, parameters $q^1\big|_{\mathcal{A}}$ were not touched by the procedure, that is why $S1\subseteq OPT(q^1)$. By analogy $\exists S2=\mathrm{O}PT(\tilde{q}^2)\subseteq OPT(q^2)$. □

Notice, that for point of minimum (q^1,q^2), satisfying (16), using the procedure in the proof, one can obtain point of minimum $(\tilde{q}^1,\tilde{q}^2)$, satisfying *complete agreement on nodes:* $\mathcal{A}(\mathrm{O}PT(\tilde{q}^1))=\mathcal{A}(\mathrm{O}PT(\tilde{q}^2))$.

Let us show, that conditions (16) of minimum (13) are in fact sufficient. We'll need the following lemma.

Lemma 4 (The uppermost optimal labeling, [8])**.** *Let $q\in supM$. Then there exists the uppermost (with respect ot operation $\sqcup$) optimal labeling $k^{\mathrm{up}}\in\mathrm{OPT}(q)$.*

Proof. From the definition of supermodular function it follows, that if $k\in\mathrm{OPT}(q)$, $k'\in\mathrm{OPT}(q)$, then $k\sqcup k'\in\mathrm{OPT}(q)$ (and else $k\sqcap k'\in\mathrm{OPT}(q)$). Taking maximum over finite set of optimal labelings we get the desired k^{up}. □

Theorem 4 (Sufficiency). *If (q^1,q^2) satisfies (16), $q^1\in supM$, $q^2\in subM$, then*

$$\Phi(q^1)+\Phi(q^2)=\min_{\Delta q}\left[\Phi(q^1+\Delta q)+\Phi(q^2-\Delta q)\right]. \tag{18}$$

Proof. We can modify q^i in order they satisfy complete agreement on nodes, which of course will not alter the value of $\sum_i\Phi(q^i)$.

$q^1\in supM$ $\Rightarrow$ there exists the uppermost optimal labeling $k^{1\mathrm{up}}\in\mathrm{OPT}(q^1)$ and else *the lowermost* optimal labeling $k^{1\mathrm{dwn}}\in\mathrm{OPT}(q^1)$.

$q^2\in subM$ $\Rightarrow$ there exists the uppermost optimal labeling $k^{2\mathrm{up}}\in\mathrm{OPT}(q^2)$, with respect to modified by Corollary 2 order in K^T, and, by analogy, $\exists k^{2\mathrm{dwn}}\in\mathrm{OPT}(q^2)$.

OPT1 and OPT2 are in *complete agreement* on $\mathcal{A}$, consequently the following equality holds:

$$\forall t \in T: \quad (k_t^{1\text{up}}, k_t^{1\text{dwn}}) = \begin{cases} (k_t^{2\text{up}}, k_t^{2\text{dwn}}), & \text{if } \mathrm{K_t} \text{ was not reversed,} \\ (k_t^{2\text{dwn}}, k_t^{2\text{up}}), & \text{if } \mathrm{K_t} \text{ was reversed .} \end{cases} \tag{19}$$

$$\text{Let } m^1 = \frac{1}{2} f(k^{1\text{up}}) + \frac{1}{2} f(k^{1\text{dwn}}) \text{ and } m^2 = \frac{1}{2} f(k^{2\text{up}}) + \frac{1}{2} f(k^{2\text{dwn}}) . \tag{20}$$

It can be verified, that $m^1 \in \mathrm{MARG}(G)$, $m^2 \in \mathrm{MARG}(G)$, and that they are node-agreed.

Booth m are convex combinations of optimal marginal vectors, corresponding to labelings $k^{1\text{up}}$, $k^{1\text{dwn}}$ and $k^{2\text{up}}$, $k^{2\text{dwn}}$, thus $\Phi(q^1) = \max_k \langle q^1, f(k) \rangle = \{q^1, m^1\}$,

$\Phi(q^2) = \max_k \langle q^2, f(k) \rangle = \langle q^2, m^2 \rangle$. Consequently $\Phi(q^1) + \Phi(q^2) = \langle q^1, m^1 \rangle + \langle q^2, m^2 \rangle =$

$\max_{(\hat{m}^1, \hat{m}^2) \in 2\mathrm{MARG}(G)} \langle q^1, \hat{m}^1 \rangle + \langle q^2, \hat{m}^2 \rangle$, which, by duality (Theorem 3), coincides with value

$\min_{\Delta q} \Phi(q^1 - \Delta q) + \Phi(q^2 + \Delta q)$. See. Fig. 1. □

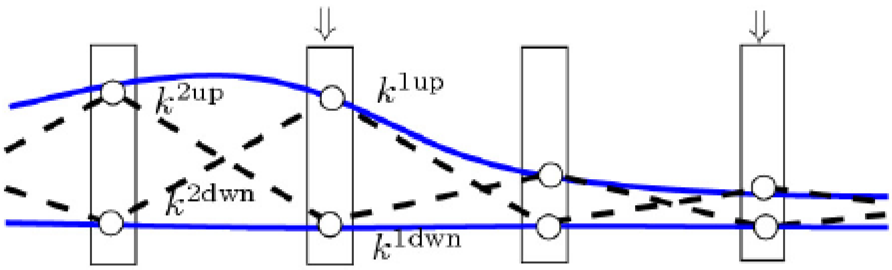

Fig. 1. A pair $(k^{1\text{up}}, k^{1\text{dwn}})$ of optimal labelings of q^1 agrees on nodes with pair $(k^{2\text{up}}, k^{2\text{dwn}})$ of optimal labelings of q^2. The problem q^2 is supermodular in the order on K^T, where objects, marked with $\Downarrow$, are reversed.

Thus, we have formulated necessary and sufficient conditions of minimum (13), which can be easily tested and allow to construct a pair of node-agreed marginal labelings.

Partial optimality. We will show, that pairs of optimal labelings we have constructed from the minimum conditions, $(k^{1\text{up}}, k^{1\text{dwn}})$, $(k^{2\text{up}}, k^{2\text{dwn}})$, allows to restrict intervals of labels in some objects of original problem, preserving the set of optimal labelings of it intact.

Lemma 5 ([2])**.** *Let* Q *be supermodular function,let* $k^* \in \operatorname{Arg}\max_k Q(k)$.

Then $\forall k \in K^T$: $Q(k \sqcap k^*) \geq Q(k)$? $Q(k \sqcup k^*) \geq Q(k)$.

Proof. We'll show the former. Labeling k^* is optimal $\Rightarrow$ $Q(k \sqcup k^*) \leq Q(k^*)$. From the supermodularity we conclude that $Q(k) + Q(k^*) \leq Q(k \sqcap k^*) + Q(k \sqcup k^*)$. Adding this two inequalities we get what was stated. □

Theorem 5 (Partial optimality)**.** *If* $q^1 \in supM$ *,* $q^2 \in subM$ *,* $q = q^1 + q^2$*,*

if $\text{OPT}(q^1)$ *and* $\text{OPT}(q^2)$ *are in complete agreement on nodes,*

then $\forall k^* \in \text{OPT}(q)$: $k^{\text{1dwn}} \leq k^* \leq k^{\text{1up}}$.

Proof. Let $Q^1(k) = \left\langle q^1, f(k) \right\rangle$, $Q^2(k) = \left\langle q^2, f(k) \right\rangle$. $Q^1 \in supM$, $Q^2 \in subM$.

Suppose for contradiction that $\exists t \in T$: $k_t^* > k_t^{\text{1up}}$, then

(1) $k^* \notin \text{OPT1} \Rightarrow Q^1((k^* \sqcap k^{\text{1up}}) \sqcup k^{\text{1dwn}}) > Q^1(k^*)$,

(2) ($k^* \notin \text{OPT1}$, OPT1 agrees on nodes with OPT2) $\Rightarrow$ $k^* \notin \text{OPT2}$ $\Rightarrow$

$Q^2((k^* \sqcap k^{\text{2up}}) \sqcup k^{\text{2dwn}}) > Q^2(k^*)$.

From the complete agreement on nodes and (19) we conclude, that:

$$(k^* \sqcap k^{\text{1up}}) \sqcup k^{\text{1dwn}} = (k^* \sqcap k^{\text{2up}}) \sqcup k^{\text{2dwn}} =: k^{**}. \tag{21}$$

And consequently $Q(k^{**}) = Q^1(k^{**}) + Q^2(k^{**}) > Q^1(k^*) + Q^2(k^*) = Q(k^*)$,

Which is in contradiction to the optimality of k^*.

Thus inequality $k_t^* \leq k_t^{\text{1up}}$ is justified, and the proof of $k_t^* \geq k_t^{\text{1dwn}}$ is analogously. See. Fig. 2. □

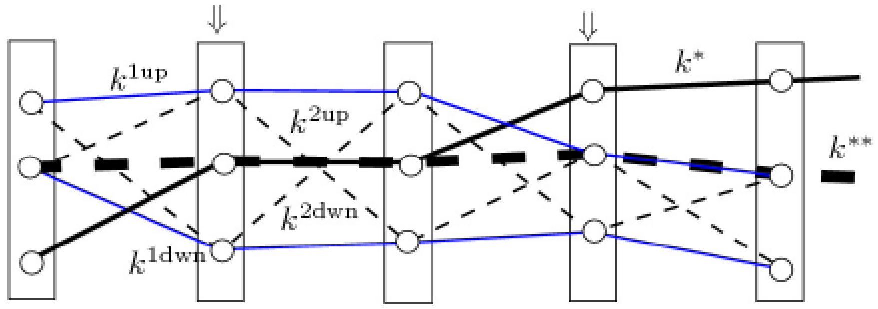

Fig 2. If for some solution k^* restrictions $k^{1\text{dwn}} \le k^* \le k^{1\text{up}}$ break down, then labeling $k^{**} := (k^* \sqcap k^{1\text{up}}) \sqcup k^{1\text{dwn}}$ appears to be strictly better, then k^*.

For some objects t it can happen, that $k_t^{1\text{up}} = k_t^{1\text{dwn}}$, i.e. in this object uppermost and lowermost labelings coincide. In this case any optimal solution of initial problem is guaranteed to pass exactly through this label. Fixation of labels in all such objects would obviously be an assured part of any optimal solution. Thus we can recover a part of optimal solution, whereas the rest part may remain an NP-hard problem.

Inclusion of solvable subclasses. We have proposed a new approach to building upper bounds on the $(max,+)$ problem, and grounded on it construction of restrictions on the set of optimal labelings / search for part of optimal labeling. The important question is a interrelation between proposed bound, and bounds proposed by other methods. In particular, we are interested in what cases partial optimality allows to specify the complete solution.

First we propose comparison to LOCAL2 relaxation in the case $K = 2$.

Theorem 6 (2MARG $\sim$ LOCAL2 when K=2). *For K=2 bounds (*14*) and (*5*) are equal:*

$$\max_{(m^1,m^2)\in 2\mathrm{MARG}(G)} \langle q^1, m^1\rangle + \langle q^2, m^2\rangle = \max_{m\in \mathrm{LOCAL2}(G)} \langle q, m\rangle, \tag{22}$$

moreover,optimal solution of one of them allows for an easy construction of optimal solution for another and visa versa.

Proof. (1) Let $(m^1, m^2) \in 2\text{MARG}$. Let, for all $e \in E$, g_e be decomposed in $g_e^1 + g_e^2$ by Lemma 2. Then for $K = 2$ there exists only one mixed second derivative c_e, and thus, either $\tilde{g}_e^1 \equiv 0$, either $\tilde{g}_e^2 \equiv 0$. Let us build $m \in \text{LOCAL2}$ as following: we put $m_e := m_e^1 \mathbf{1}_{\{c_e \geq 0\}} + m_e^2 \mathbf{1}_{\{c_e < 0\}}$, $\forall e \in E$, and $m_t := m_t^1 = m_t^2 \quad \forall t \in T$. It can be seen, that m is indeed in LOCAL2 and that $\langle q^1, m^1 \rangle + \langle q^2, m^2 \rangle = \langle q^1 + q^2, m \rangle$. Thus

$$\max_{(m^1, m^2) \in 2\text{MARG}(G)} \langle q^1, m^1 \rangle + \langle q^2, m^2 \rangle \leq \max_{m \in \text{LOCAL2}(G)} \langle q, m \rangle, \tag{23}$$

(2) Let $m \in \text{Arg} \max_{\hat{m} \in \text{LOCAL2}(G)} \langle q, \tilde{m} \rangle$. In the case $K = 2$ decomposition $q = q^1 + q^2$, $q^1 \in supM$, $q^2 \in subM$, can be selected in such a way, that m will be LOCAL2-optimal for booth problems: q^1 and q^2. Let us build two labelings $k^{1\text{up}}(m)$, and $k^{1\text{dwn}}(m)$ as following:

$$k^{1\text{up}}(m)_t := \max_{\substack{k: \\ m_{t;k} > 0}} k, \quad k^{1\text{dwn}}(m)_t := \min_{\substack{k: \\ m_{t;k} > 0}} k. \tag{24}$$

We state (follows from [9]), that $\langle q^1, m \rangle = \langle q^1, f(k^{1\text{up}}(m)) \rangle = \langle q^1, f(k^{1\text{dwn}}(m)) \rangle$.

Let us put $m^{1*} = \frac{1}{2} f(k^{1\text{up}}) + \frac{1}{2} f(k^{1\text{dwn}})$, for which it will hold $\langle q^1, m \rangle = \langle q^1, m^{1*} \rangle$.

Let us put $m^{2*} = \frac{1}{2} f(k^{2\text{up}}) + \frac{1}{2} f(k^{2\text{dwn}})$ applying the analogous considerations for q^2.

It can be verified, that $(m^{1*}, m^{2*}) \in 2\text{MARG}$, moreover $\langle q^1, m^{1*} \rangle + \langle q^2, m^{2*} \rangle = \langle q, m \rangle$. Thus

$$\max_{(m^1, m^2) \in 2\text{MARG}(G)} \langle q^1, m^1 \rangle + \langle q^2, m^2 \rangle \geq \max_{m \in \text{LOCAL2}(G)} \langle q, m \rangle. \tag{25}$$

□

From the abovestated theorem for case $K = 2$ we immediately conclude, that for solution m of LOCAL2 relaxation problem, partial optimality holds as well, i.e. if some nonzero weight m is assigned to only one of two labels in some object t, then any optimal solution of the initial $(max, +)$ problem has to pass exactly through this label.

Supermodular and submodular functions are optimized exactly by booth: LOCAL2-relaxation, and 2MARG relaxation. It can be verified, that for problems, where for each edge e, functions $g|_e$ are either super- or submodular, the abovestated theorem is else valid, and these problems appear to be tractable for booth methods.

Unfortunately, for $K > 2$ solution of LOCAL2 relaxation, in general, has nothing in common with set of optimal labelings. On the other hand for 2MARG relaxation there is trivial example with $K = 3$ labels and with two objects, for which solution of 2MARG relaxation does not allow to construct any restrictions on the set of optimal labelings. Thus, problems on trees, which are solvable exactly by LOCAL2 relaxation, forms a hard subclass for 2MARG relaxation.

We have noted, that verification of tightness of bound obtained with LOCAL2 relaxation is a hard problem. For 2MARG relaxation it is not so. Let us give without proof following statement:

Statement 2. *If $q^1 \in supM$, $q^1 \in subM$,then condition $OPT(q^1) \cap OPT(q^2) \neq \varnothing$ can be polynomially verified, and if it holds, it is possible to find a labeling $k^* \in \mathrm{OPT}(q^1) \cap \mathrm{OPT}(q^2)$.*

The proof of this statement utilizes solvability of subclass of so-called "interval" $(\vee, \wedge)$ problems, proposed in [3]. The statement itself allows constructing a new subclass of solvable $(max, +)$ problems. Thus our recommendations have the following form: search for pair (q^1, q^2) giving the tightest bound, then verifying $\mathrm{OPT}(q^1) \cap \mathrm{OPT}(q^2) \neq \varnothing$. If it holds, we found an exact solution (the problem is recognized to be in solvable subclass), otherwise we restricted the search area of optimal solution with Theorem 5.

Conclusion. We have showed, how it is possible to build upper bounds on the initial problem with it's decomposition into sum of super- and submodular. We found such decomposition, which allows to construct the tightest bound in some subclass. We showed, that from this bound restrictions of the form $k^{1\mathrm{dwn}} \leq k \leq k^{1\mathrm{up}}$ follow for all optimal labelings k^* of the initial problem.

For pattern recognition tasks, our experience tells, that if for some part of image we failed to recover optimal solution (with previously developed techniques), then on this part, a lot of approximately equivalent solutions is possible. So such approach to hard problems is very useful in practice, as for retrieval of guaranteed optimal part, as for building approximate algorithms. It can happen that proposed approach will fail to construct any restrictions on the set of optimal labelings. We show with Theorem 6, that for problems with 2 labels proposed approach is at least no worse then known once. One can observe a rother interesting fact, that several, substantially different techniques lead to one and the same upper bound.

Acknowledgments. The work was performed in the Pattern Recognition department of International Research and Scientific Center for Informational Technologies and Systems NAS and MES Ukraine. To large extend this work was motivated by research of M. Schlesinger and I. Kovtun. Author is thankful to M. Shclesinger for the problem formulation and proposed ideas, a lot of which found some place in this work. To D. Bukhantsov and I. Kovtun for conversations and verification of actual proofs.

References

[1] *Schlesinger M. I., Flach B.* Analysis of optimal labelling problems and application to image segmentation and binocular stereovision // Proceedings East-West-Vision'02 workshop / Ed. by F. Leberl, A.Ferko, New Media, 2002. —Pp. 65–81.

[2] *Schlesinger M.* Syntaksichessky analiz dvumernih zritelnih signalov v uslovijah pomeh // *Kibernetika, K.* — 1976. — T. 4. — C. 113–130.

[3] *Schlesinger M. I., Flach B.* Some solvable subclasses of structural recognition problems // Czech Pattern Recognition Workshop 2000 / Ed. by T. Svoboda; Chezh Pattern Recognition Society, February 2000 — Pp. 55–62. http://www.ki.inf.tu-dresden.de/ bflach/parts/data/sssrp.pdf .

[4] *Koval V., Schlesinger M.* Dvumernoje programmirovanije v zadachah analiza izobrageniy // *Avtomatika I telemehanika, M.*. — 1976. — T. 2. C. 149–168.

[5] *Kovaloevsky V., Koval V.*Paralelnaja vichislitelnaja schema dlja reshenija zadach dvumernogo programmirovanija // *Obrabotka i rasspoznavanije signalov,Kiev*. — 1975.—C. 38–62.

[6] *Wainwright M. J., Jaakkola T. S., Willsky A. S.* Map estimation via agreement on (hyper) trees: Message-passing and linear programming approaches: Tech. Rep. MSR-TR-2004-90: University of California, Computer Science Division

(EECS), UCB/CSD-3-1269 2003.

[7] *Kolmogorov V.* Convergent tree-reweighted message passing for energy minimization: Tech. Rep. MSR-TR-2004-90: Microsoft Research Cambridge, September 2004.

[8] *Kovtun I.* Partial optimal labeling search for a NP-hard subclass of (max, +) problems // DAGM-Symposium. — 2003. —C. 402–409.

[9] *Kovtun I. V.* Segmentacija zobragen na osnovi dostatnih umov optimalnosti v NP-povnih klasah zadach strukturnoji rozmitki: PhD / IRTC ITS NAS Ukraine, MOE Ukraine.— 2004. — C. 234.

[10] *Hammer P., Hansen P., Simeone B.* Roof duality, complementation and persistency in quadratic 0–1 optimization // *Math. Programming*. — 1984. Pp. 121–155.

[11] *Boros E., Hammer P.* Pseudo-Boolean Optimization // *Discrete Applied Mathematics* . — 2002. no. 123(1-3). Pp. 155–225.

[12] *Shekhovtsov A. Dostatochnie* uslovija optimalnosti chasti reshenija NP-polnih zadach razmetki v raspoznavaniji izobragenij. — Master's thesis, NTUU KPI, June 2005.

[13] *Hammer P.* Some network flow problems solved with pseudo-boolean programming // *Oper. Res.* — 1965. Pp. 388–399.

[14] *Hansen P.* Methods of nonlinear 0–1 programming // *Ann. Discrete Math.* — 1979. Pp. 53–70.

[15] *Greig D., Porteous B., Seheult A.* Exact maximum a posteriori estimation for binary images // *J. Royal Statistical Soc., Series B* . — 1989. Vol. 51, no. 2. Pp. 271–279.

[16] *Ishikawa H., Geiger D.* Segmentation by grouping junctions // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — 1998.

[17] *Boykov Y., Veksler O., Zabih R.* Fast approximate energy minimization via graph cuts // International Conference on Computer Vision. — Vol. 1. 1999. Pp. 377–384. http://www.csd.uwo.ca/faculty/yuri/Papers/iccv99.pdf .

[18] *Stoer M., Wagner F.* A simple min-cut algorithm // *Journal of the ACM* . —1997. Vol. 4, Pp. 585–591.

[19] *Boykov Y., Kolmogorov V.* An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision // Energy Minimization Methods in Computer Vision and Pattern Recognition. —2001. Pp. 359–374. http://www.csd.uwo.ca/faculty/yuri/Papers/pami04.pdf .

[20] *Boykov Y., Kolmogorov V.* An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision // IEEE Transactions on Pattern Analysis and Machine Intelligence. — Vol. 26. Sept 2004. Pp. 1124–1137. http://www.csd.uwo.ca/faculty/yuri/Papers/pami04.pdf .

[21] *Fisher M., Nemhauser G., Wolsey L.* An analysis of approximations for maximizing submodular set functions // *Math. Programming* . —1978. Pp. 265–294.