On Relation Between Constraint Propagation and Block-Coordinate Descent in Linear Programs*

Tomáš Dlask[0000−0002−1944−6569] (✉) and Tomáš Werner[0000−0002−6161−7157]

Faculty of Electrical Engineering, Czech Technical University in Prague
{dlaskto2,werner}@fel.cvut.cz

Abstract. Block-coordinate descent (BCD) is a popular method in large-scale optimization. Unfortunately, its fixed points are not global optima even for convex problems. A succinct characterization of convex problems optimally solvable by BCD is unknown. Focusing on linear programs, we show that BCD fixed points are identical to fixed points of another method, which uses constraint propagation to detect infeasibility of a system of linear inequalities in a primal-dual loop (a special case of this method is the Virtual Arc Consistency algorithm by Cooper et al.). This implies that BCD fixed points are global optima iff a certain propagation rule decides feasibility of a certain class of systems of linear inequalities.

Keywords: Block-Coordinate Descent · Constraint Propagation · Primal-dual Method · Linear Programming · Virtual Arc Consistency

1 Introduction

Block-coordinate descent (BCD) is a popular method in large-scale optimization which in every iteration optimizes the problem over a subset (block) of variables, keeping the remaining variables constant. Unfortunately, BCD fixed points can be arbitrarily far from global optima even for convex problems. The class of convex optimization problems for which BCD provably converges to global optima is currently quite narrow, revolving around unconstrained minimization of convex function whose non-differentiable part is separable [16].

For general (non-differentiable and/or constrained) convex problems, the set of block-optimizers in a BCD iteration can contain more than one element. It has been recently argued [22, 21] that in that case, one should choose an optimizer from the relative interior of this set. BCD updates satisfying this *relative interior rule* are not worse than any other rule to choose block-wise minimizers. Of course, this rule does not guarantee convergence to global optima.

BCD methods known as convex message passing are state-of-the art for approximately solving the dual linear programming (LP) relaxation of the MAP

* This work has been supported by the Czech Science Foundation (grant 19-09967S), the OP VVV project CZ.02.1.01/0.0/0.0/16_019/0000765, and the Grant Agency of the Czech Technical University in Prague (grant SGS19/170/OHK3/3T/13).

inference problem in graphical models [18, 8, 15] in computer vision and machine learning, which is equivalent to the Weighted (or Valued) CSP [17]. Examples are max-sum diffusion [11, 19], TRW-S [9] or MPLP [7]. These methods comply to the relative interior rule [21] (except for MPLP) and their fixed points are characterized by local consistencies (equivalent to arc consistency) of the active tuples.

Another approach to tackle the dual LP relaxation of Weighted CSP is the Virtual Arc Consistency (VAC) algorithm [2] and the similar Augmenting DAG algorithm [10, 19]. Though these are not BCD, their fixed points are also characterized by arc consistency of the active tuples. In [5] we show that this approach is related to the primal-dual method [14, §5] in linear programming and propose its generalization to any[1] linear program by replacing the arc-consistency algorithm with general constraint propagation in a system of linear inequalities.

It has been observed [4] that when BCD with the relative interior rule is applied to the dual LP relaxation of SAT, it corresponds to unit propagation. Moreover, there also exists a connection between a form of the dominating unit-clause rule and BCD with the relative interior rule applied to the dual LP relaxation of Weighted Max-SAT [4].

The above results suggest there is a close relation between BCD applied to a linear program and constraint propagation in a system of linear inequalities (and possibly equalities). In this paper we describe this relation precisely. While constraint propagation in a linear inequality system can be done in many ways, we consider the particular propagation rule that infers from a subset of inequalities that some of them are active (i.e., hold with equalities). For this rule, we show that the primal-dual approach [5] and BCD with the relative interior rule have the same fixed points. Thus, the question if a given linear program is exactly solvable by BCD can be translated to the question if feasibility of a certain system of linear inequalities is decidable by this propagation rule.

To fix notation, we consider the primal-dual pair of linear programs (LPs)

$$\begin{aligned} \max\ & c^T x & \qquad \min\ & b^T y && \text{(1a)} \\ Ax &= b & y &\in \mathbb{R}^m && \text{(1b)} \\ x &\geq 0 & A^T y &\geq c && \text{(1c)} \end{aligned}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ are constants and $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ are variables. We denote by x_i the i-th component of vector x (similarly for y, b, c), by A^j (resp. A_i) the j-th row (resp. i-th column) of A, where $i \in [n] = \{1, \dots, n\}$ and $j \in [m] = \{1, \dots, m\}$. We assume both linear programs are feasible and bounded. We assume a feasible dual solution y is given, so that $b^T y$ is an upper bound on the joint optimal value of the pair. The goal is to improve this feasible dual solution, ideally to make it dual-optimal. We further assume a finite collection $\mathcal{B} \subseteq 2^{[m]}$ of subsets (blocks) of dual variables is given.

[1] Under the assumption that an initial dual feasible solution is provided.

2 Block-Coordinate Descent with Relative Interior Rule

We start by describing BCD applied to the dual LP (1), taking into account the result [22, 21]. For convenience, we include the dual constraints into the function $f\colon \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ by defining

$$f(y) = \begin{cases} b^T y & \text{if } A^T y \geq c, \\ \infty & \text{otherwise.} \end{cases} \tag{2}$$

One BCD iteration improves a feasible dual solution y by choosing a block $B \in \mathcal{B}$ and optimizing over variables $y_B = (y_i)_{i \in B}$, keeping the remaining variables $y_{-B} = (y_i)_{i \in [m]-B}$ constant. That is, it changes y_B to satisfy

$$y_B \in \operatorname*{argmin}_{y'_B \in \mathbb{R}^B} f(y'_B, y_{-B}). \tag{3}$$

The set $\operatorname{argmin}_{y'_B \in \mathbb{R}^B} f(y'_B, y_{-B}) \subseteq \mathbb{R}^B$ of block-wise minimizers is a non-empty convex polyhedron. If this polyhedron contains more than one point, we need to choose a single element from this polyhedron. To satisfy the relative interior rule, the update must be modified to

$$y_B \in \operatorname{ri} \operatorname*{argmin}_{y'_B \in \mathbb{R}^B} f(y'_B, y_{-B}) \tag{4}$$

where $\operatorname{ri} X$ denotes the relative interior of a convex set X [12, §2.1]. The following are the main results of [22, 21]:

Definition 1. *A point y feasible to the dual in* (1) *is*

- *a local minimum (LM) of f w.r.t. $\mathcal{B}$ if* (3) *holds for all $B \in \mathcal{B}$,*
- *an interior local minimum (ILM) of f w.r.t. $\mathcal{B}$ if* (4) *holds for all $B \in \mathcal{B}$,*
- *a pre-interior local minimum (pre-ILM) of f w.r.t. $\mathcal{B}$ if there is an ILM y' such that y is in a face of the polyhedron $\{ y \mid A^T y \geq c \}$ containing y' in its relative interior.*

Theorem 1. *Let $(B_i)_{i=1}^\infty$ be a sequence of blocks $B_i \in \mathcal{B}$ that contains each element of $\mathcal{B}$ an infinite number of times. Let $(y^i)_{i=1}^\infty$ be a sequence produced by the BCD method, where the blocks are visited in the order given by $(B_i)_{i=1}^\infty$.*

A. If $(y^i)_{i=1}^\infty$ satisfies (4) *and y^1 is an ILM, then y^i is an ILM for all i.*
B. If $(y^i)_{i=1}^\infty$ satisfies (4) *and y^1 is a pre-ILM, then y^i is an ILM for some i.*
C. If $(y^i)_{i=1}^\infty$ satisfies (3) *and y^1 is a pre-ILM, then $b^T y^i = b^T y^1$ for all i.*
D. If $(y^i)_{i=1}^\infty$ satisfies (4) *and y^1 is not a pre-ILM, then $b^T y^i < b^T y^1$ for some i.*

Thus, when we are at a pre-ILM, the objective cannot be improved by any further BCD iterations. When we are not at a pre-ILM, BCD with the relative interior rule inevitably improves the objective in a finite number of iterations.

3 Primal-Dual Approach

Let us now focus on the second of the two approaches we consider in this paper.

By the *complementary slackness* theorem [14, 13], a primal feasible solution x and a dual feasible solution y to (1) are optimal if and only if $x_i(A_i^T y - c) = 0$ for all $i \in [n]$. In addition, x^* is in the relative interior of the primal optimizers and y^* is in the relative interior of the dual optimizers if and only if they satisfy *strict complementary slackness* condition [23] $(x_i^* = 0) \oplus (A_i^T y^* = c)$ for all $i \in [n]$ where $\oplus$ denotes exclusive disjunction. If both primal and dual are feasible and bounded, there always exist such x^*, y^* [12, Theorem 2.1.3].

The iteration of the primal-dual approach proceeds as follows. Denoting

$$K(y) = \{i \in [n] \mid A_i^T y = c_i\}, \tag{5}$$

the index set of dual constraints active at y, the complementary slackness condition reads

$$Ax = b \tag{6a}$$
$$x_i \geq 0 \qquad \forall i \in K(y) \tag{6b}$$
$$x_i = 0 \qquad \forall i \in [n] - K(y) \tag{6c}$$

Thus, y is dual-optimal for (1) if and only if system (6) is feasible. By Farkas' lemma [13, §6] (or by LP duality), (6) is infeasible if and only if the system

$$b^T \bar{y} < 0 \tag{7a}$$
$$A_i^T \bar{y} \geq 0 \qquad \forall i \in K(y) \tag{7b}$$

is feasible. In that case, any solution $\bar{y}$ to (7) is an improving direction for the dual (1) from point y, i.e., there is $\epsilon > 0$ such that $b^T(y + \epsilon\bar{y}) < b^T y$ and $A^T(y + \epsilon\bar{y}) \geq c$. Updating $y \leftarrow y + \epsilon\bar{y}$ yields a better dual feasible solution.

The described approach is similar to the well-known primal-dual method [14, §5], where complementary slackness (6) is not required strictly but only its violation is minimized. The motivation for the method is that problem (6) may be easier to solve than (1), possibly by combinatorial algorithms [14, §6].

3.1 Constraint Propagation

Deciding feasibility of a system of linear inequalities (such as (6)) can be too costly for large instances. Therefore, we proposed in [5, §2] to do it by constraint propagation: using a small fixed set of inference (or propagation) rules, we iteratively infer new linear inequalities from the system and add them to the system. If a contradictory inequality is inferred, the initial system was infeasible; then an infeasibility certificate (such as $\bar{y}$ in (7)) is constructed from the propagation history. The drawback of this method is that it is in general refutation-incomplete: it may not infer a contradiction even if the system is infeasible.

While in [5, §2] we did not restrict the form of the used inference rules, here we consider one particular form: *choose a subset of the inequalities and infer* *which of them are active (i.e., hold with equality)*. For the particular case of system (6), this means we choose a subset $B \in \mathcal{B}$ of equalities (6a) and decide if they, together with (6b) and (6c), imply that some of the inequalities (6b) is active, i.e., $x_i = 0$. Indeed, this can be seen as inferring the inequality $x_i \leq 0$ from the system. It is our key observation in this paper that with this propagation rule the primal-dual approach has the same fixed points as BCD (see §4).

Precisely, the algorithm first initializes $K = K(y)$ and then repeats the following iteration: choose $B \in \mathcal{B}$, find all indices $i \in K$ for which the system

$$
\begin{aligned}
A^j x &= b_j & \forall j &\in B && \text{(8a)}\\
x_i &\geq 0 & \forall i &\in K && \text{(8b)}\\
x_i &= 0 & \forall i &\in [n] - K && \text{(8c)}
\end{aligned}
$$

implies[2] $x_i = 0$, and remove these indices from K. If the set K shrinks so much that system (8) becomes infeasible for some $B \in \mathcal{B}$, then clearly the original system (6) is infeasible. Next, we analyze this algorithm, showing that its properties are analogous to the well-known arc-consistency algorithm.

Definition 2. *For $B \subseteq [m]$, a set $K \subseteq [n]$ is B-consistent if for every $i \in K$ system* (8) *does not imply $x_i = 0$, i.e., if the system*

$$
\begin{aligned}
A^j x &= b_j & \forall j &\in B && \text{(9a)}\\
x_i &> 0 & \forall i &\in K && \text{(9b)}\\
x_i &= 0 & \forall i &\in [n] - K && \text{(9c)}
\end{aligned}
$$

is feasible. For $\mathcal{B} \subseteq 2^{[m]}$, K is $\mathcal{B}$-consistent if it is B-consistent for all $B \in \mathcal{B}$.

Proposition 1. *If K and K' are $\mathcal{B}$-consistent, then $K \cup K'$ is $\mathcal{B}$-consistent.*

Proof. If (9) for some $B \in \mathcal{B}$ is satisfied by x (resp. x') for K (resp. K'), then it is satisfied by $(x + x')/2$ for $K \cup K'$. $\square$

By Proposition 1, the $\mathcal{B}$-consistent sets form a join-semilattice w.r.t. the inclusion. Therefore, for any $K \subseteq [n]$, either there is no $\mathcal{B}$-consistent subset of K or there exists the unique maximal $\mathcal{B}$-consistent subset of K.

Definition 3. *The* propagator *over block $B \subseteq [m]$ is the map $P_B \colon 2^{[n]} \cup \{\perp\} \to 2^{[n]} \cup \{\perp\}$ defined by*[3]*:*

[2] By saying that (8) implies $x_i = 0$ we mean that $x_i = 0$ holds for all x satisfying (8). This can be decided by, e.g., projecting polyhedron (8) onto the i-th coordinate. The projection is a singleton set $\{0\}$ if and only if (8) implies $x_i = 0$. The projection can be computed by the Fourier-Motzkin elimination or by maximizing x_i subject to (8) (which equals 0 if and only if (8) implies $x_i = 0$).

[3] Note that $P_B(K) = \perp$ is different from $P_B(K) = \emptyset$, since system (8) can be feasible even for $K = \emptyset$ (if $b = 0$).

- *If* $K = \bot$, *then* $P_B(K) = \bot$.
- *If* $K \subseteq [n]$ *and* (8) *is infeasible, then* $P_B(K) = \bot$.
- *If* $K \subseteq [n]$ *and* (8) *is feasible, then* $P_B(K) \subseteq [n]$ *and* $i \in P_B(K)$ *if and only if* (8) *does not imply* $x_i = 0$.

Clearly, a set $K \subseteq [n]$ is B-consistent if and only if $P_B(K) = K$, and K is $\mathcal{B}$-consistent if and only if $P_B(K) = K$ for all $B \in \mathcal{B}$.

Proposition 2. *Map* $P_B(\cdot)$ *satisfies the axioms of a closure operator unless*[4] $P_B(\cdot) = \bot$, *i.e., for all* $K, K' \subseteq [n]$ *such that* $P_B(K), P_B(K') \neq \bot$ *we have*

- $P_B(P_B(K)) = P_B(K)$ *(idempotence)*
- $P_B(K) \subseteq K$ *(intensivity)*
- $K' \subseteq K \implies P_B(K') \subseteq P_B(K)$ *(monotonicity).*

Proof. Idempotence and intensivity are straightforward. To prove monotonicity, let $K' \subseteq K$ and let H' (resp. H) be the polyhedron defined by (8) for K' (resp. K). Clearly, $\emptyset \neq H' \subseteq H$. If $i \in [n] - P_B(K)$, the projection of H onto x_i contains only 0. Therefore, the projection of H' onto x_i also contains only 0, i.e., (8) for K' implies $x_i = 0$, hence $i \in [n] - P_B(K')$. Thus $P_B(K') \subseteq P_B(K)$. □

Definition 4. *Given* $K \in 2^{[n]} \cup \{\bot\}$, *the* propagation algorithm *repeats the following iteration: find* $B \in \mathcal{B}$ *such that* $P_B(K) \neq K$ *and set* $K \leftarrow P_B(K)$. *If no such* $B \in \mathcal{B}$ *exists, return the final* K.

The propagation algorithm terminates in a finite number of steps. If at any iteration we get $K = \bot$, the algorithm terminates due to $P_B(\bot) = \bot$ for all $B \in \mathcal{B}$. Otherwise, by intensivity of $P_B(\cdot)$, K can decrease only a finite number of times.

Proposition 3. *If* $K \in [n]$ *has a* $\mathcal{B}$*-consistent subset, the propagation algorithm returns the maximal* $\mathcal{B}$*-consistent subset of* K.

Proof. The propagation algorithm creates a finite decreasing chain $K_1 \supset K_2 \supset K_3 \supset \cdots$ where $K_1 = K$ and $K_{l+1} = P_{B_l}(K_l)$ where $B_l \in \mathcal{B}$ is the block chosen in the l-th step. Let L be arbitrary $\mathcal{B}$-consistent subset of K. We will prove by induction that $L \subseteq K_l$ for all l. Clearly, $L \subseteq K = K_1$. If $L \subseteq K_l$, then

$$L = P_{B_l}(L) \subseteq P_{B_l}(K_l) = K_{l+1} \tag{10}$$

where the first equality follows from $\mathcal{B}$-consistency of L and the inclusion follows from monotonicity of $P_{B_l}(\cdot)$ by Proposition 2. See that it cannot happen that $P_{B_l}(K_l) = \bot$ for any l because (8) is feasible for all $B \in \mathcal{B}$ for L and $L \subseteq K_l$. □

By Proposition 3, the result of the propagation algorithm does not depend on the order in which the elements of $\mathcal{B}$ are visited. Thus, we can introduce the operator $P_{\mathcal{B}}$: $2^{[n]} \cup \{\bot\} \to 2^{[n]} \cup \{\bot\}$ where $P_{\mathcal{B}}(K)$ is the unique result of the algorithm with input K.

[4] The exception of $\bot$ could be removed by augmenting the set $2^{[n]}$, partially ordered by set inclusion, with $\bot$ as its least element.

Proposition 4. *The operator $P_{\mathcal{B}}(\cdot)$ satisfies the axioms of a closure operator unless $P_{\mathcal{B}}(\cdot) = \bot$.*

Proof. Idempotence and intensivity follow directly from the proof of Proposition 3. We will prove monotonicity by contradiction: let $K' \subseteq K$ and $P_{\mathcal{B}}(K) \subset P_{\mathcal{B}}(K')$. By intensivity, we obtain $P_{\mathcal{B}}(K') \subseteq K' \subseteq K$. However, $P_{\mathcal{B}}(K)$ is $\mathcal{B}$-consistent and $P_{\mathcal{B}}(K') \subseteq K$. Since $P_{\mathcal{B}}(K) \subset P_{\mathcal{B}}(K')$, $P_{\mathcal{B}}(K)$ is not the maximal $\mathcal{B}$-consistent subset of K, which is contradictory with Proposition 3. $\square$

Due to Proposition 4, $P_{\mathcal{B}}(K)$ can be called the *$\mathcal{B}$-consistency closure* of K. Observe that the properties of $P_B(\cdot)$ and $P_{\mathcal{B}}(\cdot)$ are analogous to the properties of the arc-consistency propagator and arc-consistency closure, respectively. In more general view, the propagator resembles domain-based constraint propagation [1], where stability under union corresponds to the property given by Proposition 1 and Φ-closure corresponds to $\mathcal{B}$-consistent closure.

If $P_{\mathcal{B}}(K(y)) = \bot$, then system (6) is infeasible. Then there exists an improving direction $\bar{y}$ satisfying (7). Such an improving direction can be constructed from the history of the propagation, as we describe in Appendix B. But note that improving directions are not necessary for our analysis in this paper as we only consider the fixed points of the primal-dual approach.

As propagation is not refutation-complete, $P_{\mathcal{B}}(K(y)) \neq \bot$ does not in general imply that (6) is feasible. Consequently, $b^T y$ is not the optimal value of the pair (1) but only its upper bound.

The primal-dual approach with the described propagation is used, under various names, in several existing methods. One example is the VAC algorithm [2] and the Augmenting DAG algorithm [10, 19], where the primal problem (1) is the basic LP relaxation of the Weighted CSP and our propagation is equivalent to the arc-consistency algorithm [5]. An approach proposed in [5] to upper-bound the LP relaxation of Weighted Max-SAT is (up to technical details) another example. If the minimization of a convex piecewise-affine function is expressed as an LP, then our method subsumes the sign relaxation technique introduced in [20] and further developed in [3].

4 Relation Between the Approaches

We now state the relation between BCD with the relative interior rule (§2) and the primal-dual approach in which system (6) is solved by constraint propagation as described in §3.1. The proof of the theorem is in Appendix A.

Theorem 2. *Let y be a feasible point for dual* (1)*. Then:*

- *y is an LM of dual* (1) *w.r.t. $\mathcal{B}$ if and only if $P_B(K(y)) \neq \bot$ for all $B \in \mathcal{B}$,*
- *y is an ILM of dual* (1) *w.r.t. $\mathcal{B}$ if and only if $P_{\mathcal{B}}(K(y)) = K(y)$,*
- *y is a pre-ILM of dual* (1) *w.r.t. $\mathcal{B}$ if and only if $P_{\mathcal{B}}(K(y)) \neq \bot$.*

Theorem 2 characterizes the previously introduced types of local minima in BCD by local consistency conditions. It also shows that BCD with relative interior rule cannot improve the fixed points of the primal-dual approach based on propagation and vice versa. This yields the following corollary.

Corollary 1. *The following statements are equivalent:*

- *For all feasible y for the dual* (1)*, if* (6) *is infeasible then $P_{\mathcal{B}}(K(y)) = \bot$ (i.e., propagation is refutation-complete).*
- *Any ILM y of the dual* (1) *w.r.t. $\mathcal{B}$ is a global optimum.*

This result shows that the question whether BCD fixed points are global minima for a given class of LPs can be reformulated as the question whether constraint propagation decides feasibility of a certain class of linear inequalities.

5 Other Forms of Linear Programs

It is well-known that linear programs come in different forms [14, §2.1] which can be easily transformed to each other, preserving global optima. One can ask if the propagation algorithm can be formulated and the equivalence with BCD holds also for different forms than (1). This question is non-trivial because transformations that preserve global optima do not necessarily preserve (pre-)interior local optima [6]. We show that independently of the formulation, if we use the propagation rule that infers activity of inequality constraints (as we mentioned in the beginning of §3.1), the two approaches remain equivalent.

5.1 Primal LP with Inequalities and Non-Negative Variables

Consider for example the primal-dual pair

$$
\begin{aligned}
\max c^T x &\qquad\qquad& \min b^T y &\qquad (11a)\\
Ax \le b &\qquad\qquad& y \ge 0 &\qquad (11b)\\
x \ge 0 &\qquad\qquad& A^T y \ge c &\qquad (11c)
\end{aligned}
$$

that can be equivalently reformulated [13] by introducing slack variables $s_j \ge 0$, $j \in [m]$ as

$$
\begin{aligned}
\max c^T x &\qquad\qquad& \min b^T y &\qquad (12a)\\
Ax + s = b &\qquad\qquad& y \in \mathbb{R}^m &\qquad (12b)\\
x \ge 0 &\qquad\qquad& A^T y \ge c &\qquad (12c)\\
s \ge 0 &\qquad\qquad& y \ge 0 &\qquad (12d)
\end{aligned}
$$

which is in the form (1). See that BCD in the duals (11) and (12) is identical.

The propagation rules presented previously in §3.1 for the LP (12) correspond to deciding which s_j and x_i are forced to be zero. Clearly, setting $s_j = 0$ corresponds to setting $A^j x = b_j$ and enforcing $s_j > 0$ implies $A^j x < b_j$. Thus, instead of rewriting (11) into (12), we can apply propagation directly on the primal (11) except that when considering the system (9) for some $B \in \mathcal{B}$, we will instead of a single set K use two sets $K^X \subseteq [n]$ and $K^S \subseteq [m]$ that indicate which inequalities need to be satisfied strictly and which with equality, i.e., we will use

$$A^j x < b_j \quad \forall j \in K^S \cap B \qquad x_i > 0 \quad \forall i \in K^X \tag{13a}$$
$$A^j x = b_j \quad \forall j \in ([m] - K^S) \cap B \qquad x_i = 0 \quad \forall i \in [n] - K^X \tag{13b}$$

instead of (9). Clearly, deciding which inequalities among $Ax \leq b$ in primal (11) need to be satisfied with strict inequality (resp. with equality) by considering a set $K^S \subseteq [m]$ is in one-to-one correspondence with deciding which slack variables s_j in (12) can be non-zero (resp. are forced to be zero).

5.2 Primal LP with Inequalities and Unconstrained Variables

Another general primal-dual pair that we are going to consider is

$$\max c^T x \qquad \min b^T y \tag{14a}$$
$$Ax \leq b \qquad y \geq 0 \tag{14b}$$
$$x \in \mathbb{R}^n \qquad A^T y = c \tag{14c}$$

where y is optimal for the dual if and only if there exists $x \in \mathbb{R}^n$ such that

$$A^j x \leq b_j \qquad \forall j \in K'(y) \tag{15a}$$
$$A^j x = b_j \qquad \forall j \in [m] - K'(y) \tag{15b}$$

where $K'(y) = \{j \in [m] \mid y_j = 0\}$ and (15) again follows from complementary slackness. From this point, we could completely repeat the reasoning in §3.1 and prove the same theorem as in §4 except that we would replace $K(y)$ by $K'(y)$, replace condition $A_i^T y > c$ (resp. $A_i^T y = c$) by $y_j > 0$ (resp. $y_j = 0$) and infer whether the inequality $A^j x \leq b_j$ should hold strictly or with equality instead of inferring it for $x_i \geq 0$. This is based on similarity between (15) and (6).

5.3 Redundant Constraints

It was observed in [6] that adding redundant constraints into an LP has significant influence on its solvability by (block-)coordinate descent. Using our results from this paper, we are able to explain this quite naturally.

As an example, consider the following LP relaxation of weighted vertex cover on a graph (V, E) with vertex weights $w : V \to \mathbb{R}^+$ together with its dual

$$\min w^T x \qquad \max \textstyle\sum_{\{i,j\} \in E} y_{ij} \tag{16a}$$
$$x_i + x_j \geq 1 \qquad y_{ij} \geq 0 \qquad \forall \{i,j\} \in E \tag{16b}$$
$$x_i \geq 0 \qquad \textstyle\sum_{j \in N_i} y_{ij} \leq w_i \qquad \forall i \in V \tag{16c}$$

where N_i is the set of neighbors of vertex i in the graph. If we optimized the primal or the dual (16) by coordinate descent along individual variables (i.e., blocks of size 1), there are interior local optima[5] that are not global optima [6]. However, if we add redundant constraints $x \le 1$ to the primal, we obtain

$$\begin{aligned}
&\min w^T x && \max \textstyle\sum_{\{i,j\}\in E} y_{ij} + \sum_{i\in V} z_i && && \text{(17a)}\\
&x_i + x_j \ge 1 && y_{ij} \ge 0 && \forall \{i,j\} \in E && \text{(17b)}\\
&x_i \ge 0 && z_i + \textstyle\sum_{j\in N_i} y_{ij} \le w_i && \forall i \in V && \text{(17c)}\\
&x_i \le 1 && z_i \le 0 && \forall i \in V. && \text{(17d)}
\end{aligned}$$

By the result in [6], any interior local optimum of dual (17) w.r.t. blocks[6] consisting of variables y_{ij}, z_i, and z_j for each $\{i,j\} \in E$ is a global optimum.

The explanation for the difference between (non-)optimality for the different formulations lies in the fact that in case of (16), we can only propagate equality in constraints (16b) and $x_i = 0$. However, in (17), we are also able to propagate $x_i = 1$ due to the added constraint $x_i \le 1$. This results in a stronger propagation algorithm which is even refutation-complete for this case.

This also holds for the LP formulation of min-*st*-cut and its dual, maximum flow, which was also considered in [6, §4.3]. Adding redundant bounds $0 \le x_i \le 1$ for variables in min-*st*-cut results in optimality of BCD on its dual. However, the dual of the usual formulation of min-*st*-cut (i.e., without these bounds) is not amenable to BCD [6, §4.3]. This difference is now explained by the possibility of the underlying propagation algorithm to set these variables to their bounds, i.e., set $x_i = 0$ or $x_i = 1$ which is not possible if variables x are unbounded.

The result in this paper therefore also sheds light on which constraints are useful in terms of propagation or BCD even though they are redundant from the point of global optimality of the original linear program.

6 Conclusion

Even though propagation in a system of linear inequalities can be performed in many ways, we have defined a propagation algorithm which not only has natural and useful properties, but it also allows full characterization of types of local minima in BCD. Additionally, there is a tight connection between the fixed points of BCD with relative interior rule and the fixed points of primal-dual approach based on this propagation algorithm. Despite the fact that both

[5] In case of the dual, we maximize, so we should talk about interior local maxima, but this relation is straightforward by inverting the sign in the criterion and changing maximization to minimization.

[6] In analogy with [6, §3 equation (7)], $z_i = \min\{w_i - \sum_{j\in N_i} y_{ij}, 0\}$ $\forall i \in V$ holds in any optimal solution of dual (17) and so the dual can be equivalently reformulated as maximization of a concave piecewise-affine function with non-negative variables, which makes optimization along these blocks simpler. In detail, variables z were eliminated and thus we update only each y_{ij} separately.

algorithms may not reach a global optimum, none of the algorithms can improve the fixed points of the other.

We argued that the propagation algorithm can be generalized to linear programs in any form. In detail, BCD in the dual for a given set of blocks $\mathcal{B}$ corresponds to propagating which primal constraints given by complementary slackness should be active and which inactive while inferring only from subsets of the constraints given by sets in $\mathcal{B}$.

We believe that our findings are interesting for the theory of BCD as they explain what kind of local consistency is reached by any BCD algorithm (both with or without relative interior rule) on any LP. E.g., As shown in [21], since both TRW-S [9] and max-sum diffusion [11, 19] satisfy the relative interior rule, their fixed point conditions are equivalent to the proposed local consistency condition if applied to the specific LP formulations which these algorithms optimize.

This tight connection between the decidability of feasibility of a system of linear inequalities by refutation-incomplete propagation and BCD may provide theoretical ground for analysis of BCD in terms of constraint propagation. Moreover, it may result in newly discovered classes of problems optimally solvable by BCD or better design for choices of blocks of variables so that the propagation is more effective and BCD may reach better local optima. This connection also precisely explains the differences in applicability of BCD caused by minor changes in the formulation of the optimized LP, as discussed in §5.3.

The practical impact of these results is mainly focused on approximately optimizing challenging large-scale LPs which are not solvable by off-the-shelf LP solvers due to their super-linear space complexity. Propagation algorithms subsumed (up to technical details) by the proposed one were previously derived ad-hoc for specific LPs [2, 5, 10, 20, 3] where they provided useful solutions which were often close to global optima. Presenting all of these algorithms in a single framework may simplify design of similar algorithms in the future.

A Proofs

Proposition 5. *Let y be feasible for the dual* (1) *and let $B \subseteq [m]$. Block of variables y_B satisfies* (4) *if and only if $P_B(K(y)) = K(y)$.*

Proof. For the 'only-if' direction, construct the dual (1) restricted only to the variables y_B as follows:

$$\begin{aligned} &\max k^T x && \min \textstyle\sum_{j\in B} b_j y_j && && \text{(18a)}\\ &A^j x = b_j && y_j \in \mathbb{R} && \forall j \in B && \text{(18b)}\\ &x_i \ge 0 && \textstyle\sum_{j\in B} A_{ji} y_j \ge k_i && \forall i \in [n] && \text{(18c)} \end{aligned}$$

where $k_i = c_i - \sum_{j\in[m]-B} A_{ji} y_j$ are viewed as constants determined by the remaining variables that are not in the block and A_{ji} is the entry of matrix A on j-th row and i-th column. The problem on left is the corresponding primal.

Since y_B is in the relative interior of optimizers of the dual (18) by our assumption, there must exist a solution $x \in \mathbb{R}^n_+$ for the primal (18) such that strict complementary slackness holds. The condition for this case reads

$$\textstyle\sum_{j\in B} A_{ji} y_j = k_i \iff x_i > 0 \qquad \forall i \in [n], \tag{19}$$

therefore x satisfies $x_i = 0$ $\forall i \in [n] - K(y)$ and $x_i > 0$ $\forall i \in K(y)$ by definition of $K(y)$ and k_i. By feasibility of x for primal (18), we have that $i \in P_B(K(y))$ for all $i \in K(y)$. By intensivity of $P_B(\cdot)$, we obtain $P_B(K(y)) = K(y)$.

For the 'if' direction, assume $P_B(K(y)) = K(y)$, then there must exist a solution $x \in \mathbb{R}^n$ for (9) where $K = K(y)$. This vector x is a feasible solution for the primal (18). By definition of $K(y)$ in (5) and definition of k_i, it follows that strict complementary slackness (19) is satisfied in (18), therefore both x and y_B lie in the relative interior of optimizers of the primal-dual pair (18). □

Corollary 2. *Let y be feasible for dual* (1)*. Then y is an ILM of dual* (1) *w.r.t. $\mathcal{B}$ if and only if $P_{\mathcal{B}}(K(y)) = K(y)$.*

Proof. By definition, y is an ILM of dual (1) w.r.t. $\mathcal{B}$ if (4) holds $\forall B \in \mathcal{B}$. Applying Proposition 5, this is equivalent to $P_B(K(y)) = K(y)$ $\forall B \in \mathcal{B}$, i.e., $P_{\mathcal{B}}(K(y)) = K(y)$. □

Proposition 6. *Let y be feasible for the dual* (1) *and let $B \subseteq [m]$. Block of variables y_B satisfies* (3) *if and only if $P_B(K(y)) \neq \bot$.*

Proof. Block y_B satisfies (3) if and only if it is optimal for the dual (18), which happens if and only if there exists $x \in \mathbb{R}^n_+$ satisfying complementary slackness. By definition of $K(y)$, complementary slackness conditions are equivalent to (8) for $K = K(y)$ which is feasible if and only if $P_B(K(y)) \neq \bot$. □

Proposition 7. *If point x is in the relative interior of optimizers of the primal* (1)*, then the set $\{i \in [n] \mid x_i = 0\}$ is minimal w.r.t. inclusion among all optimal solutions and is unique.*

Proof. By contradiction: let x (resp. y) be from the relative interior of optimizers of primal (resp. dual) (1). Let x' be also optimal for the primal and let $\{i \in [n] \mid x'_i = 0\}$ be smaller and/or different. Then, there is $k \in [n]$ such that $x'_k > 0$ and $x_k = 0$. Since x and y are in the relative interior of optimizers, they satisfy strict complementary slackness, thus $A_k^T y > c_k$. Complementary slackness is satisfied by all pairs of primal and dual optimal solutions, but x' and y do not satisfy it because $x'_k > 0$ and $A_k^T y > c_k$, hence x' is not optimal. □

Proposition 8. *Let y be a feasible point for dual* (1) *and let $B \subseteq [m]$ so that $P_B(K(y)) = K \neq \bot$. Then, there exists a feasible point y' such that $b^T y = b^T y'$ and $P_B(K(y')) = K(y') = K$.*

Proof. Consider the primal-dual pair

$$\begin{aligned}
&\max 0 && \min b^T\bar{y} && && \text{(20a)}\\
&A^j x = b_j && \bar{y}_j \in \mathbb{R} && \forall j \in B && \text{(20b)}\\
&x_i = 0 && - && \forall i \in [n] - K(y) && \text{(20c)}\\
&x_i \geq 0 && A_i^T\bar{y} \geq 0 && \forall i \in K(y) && \text{(20d)}\\
&- && \bar{y}_j = 0 && \forall j \in [m] - B. && \text{(20e)}
\end{aligned}$$

which was simplified in the sense that if some primal (resp. dual) variable equals zero, then we can omit the corresponding dual (resp. primal) constraint without changing the problem since the whole column (resp. row) of A can be set to zero.

Let x (resp. $\bar{y}$) be in the relative interior of optimizers for the primal (resp. dual) (20). By Proposition 7 applied on matrix A with only rows in B and only columns in $K(y)$, if some $x_i = 0$, then this is the only value x_i can take, therefore $i \notin K$ because primal (20) is (8) for $K(y)$. If some variable x_i can take a non-zero value in (20), then it is non-zero again by Proposition 7 and $i \in K$.

Since the pair of optimal solutions $x, \bar{y}$ lies in the relative interior of optimizers, they satisfy strict complementary slackness in this form:

$$x_i = 0 \wedge A_i^T\bar{y} > 0 \qquad \forall i \in K(y) - K \tag{21a}$$

$$x_i > 0 \wedge A_i^T\bar{y} = 0 \qquad \forall i \in K. \tag{21b}$$

We will now choose any ϵ such that

$$0 < \epsilon < \frac{c_i - A_i^T y}{A_i^T\bar{y}} \qquad \forall i \in [n] - K(y) \text{ such that } A_i^T\bar{y} < 0 \tag{22}$$

where the upper bound is positive because $A_i^T\bar{y} < 0$ by the condition in the upper bound and for all $i \in [n] - K(y)$, $c_i - A_i^T y < 0$ by feasibility of y and definition of $K(y)$. Therefore, $(c_i - A_i^T y)/(A_i^T\bar{y})$ is positive for all i considered in (22) and there exists some ϵ satisfying (22). We choose any ϵ satisfying (22) and claim that $y' = y + \epsilon\bar{y}$ satisfies the required conditions.

- If $i \in K(y) - K$, then $A_i^T\bar{y} > 0$ by (21) and $A_i^T y = c_i$ by definition of $K(y)$. Therefore, $A_i^T y' = c_i + \epsilon A_i^T\bar{y} > c_i$, so $i \notin K(y')$, i.e., $i \in [n] - K(y')$.
- If $i \in K$, then $A_i^T y = c_i$ because $i \in K = P_B(K(y)) \subseteq K(y)$ and $A_i^T\bar{y} = 0$ by (21). Therefore, $A_i^T y' = c_i + \epsilon \cdot 0 = c_i$ and $i \in K(y')$.
- If $i \in [n] - K(y)$, then $A_i^T y > c_i$ by definition of $K(y)$ and $A_i^T\bar{y}$ can have any sign. We distinguish the following cases:
  - If $A_i^T\bar{y} \geq 0$, then $A_i^T y' > c_i + \epsilon A_i^T\bar{y} \geq c_i$, so $i \notin K(y')$, i.e., $i \in [n] - K(y')$.
  - If $A_i^T\bar{y} < 0$, then by definition of ϵ, $\epsilon < (c_i - A_i^T y)/(A_i^T\bar{y})$ which implies $A_i^T y' = A_i^T y + \epsilon A_i^T\bar{y} > c_i$, hence $i \notin K(y')$, i.e., $i \in [n] - K(y')$.

Therefore, $[n] - K(y) \subseteq [n] - K(y')$ and $K(y) - K \subseteq [n] - K(y')$, which results in $[n] - K \subseteq [n] - K(y')$. This together with $K \subseteq K(y')$ yields $K(y') = K$.

Point y' is feasible since $A_i^T y' \geq c_i$ for all $i \in [n]$ as just shown above. Because the optimal value of the primal (20) is 0 and $\bar{y}$ is an optimal dual solution, it follows from strong duality that $b^T\bar{y} = 0$, therefore $b^T y' = b^T y + \epsilon b^T \bar{y} = b^T y$. By idempotency of $P_B(\cdot)$, it follows that $P_B(K) = K$, i.e., $P_B(K(y')) = K(y')$. $\square$

Remark 1. The point y' constructed in Proposition 8 can in fact be obtained by updating block y_B to satisfy (4). By construction of y' from y derived in Proposition 8, $y_j = y'_j \ \forall j \in [m] - B$, therefore only the variables in block B change. Combining Proposition 5 with $P_B(K(y')) = K(y')$ implies that block y'_B is in the relative interior of optimizers of the dual (1) restricted to this block.

Proposition 9. *Let y be a feasible point for dual* (1) *such that $P_{\mathcal{B}}(K(y)) \neq \perp$, then y is a pre-ILM of dual* (1) *w.r.t. $\mathcal{B}$.*

Proof. By the definition of $P_{\mathcal{B}}(\cdot)$, there must exist a finite sequence $(B_l)_{l=1}^{L}$ for $B_l \in \mathcal{B}, l \in [L]$ such that

$$P_{B_L}(P_{B_{L-1}}(P_{B_{L-2}}(\cdots P_{B_2}(P_{B_1}(K(y)))\cdots))) = K \tag{23}$$

and $P_{\mathcal{B}}(K) = K$. In other words, the sequence corresponds to the order of the blocks B applied in the propagation algorithm until a fixed point is reached.

We construct a sequence $(y^l)_{l=1}^{L+1}$, $y^l \in \mathbb{R}^m$ where $y^1 = y$ and y^{l+1} is constructed from y^l as in the proof of Proposition 8 for $B = B_l$. By induction and properties of the construction, since y^1 is feasible, the other points $y^2, y^3, ..., y^{L+1}$ are also feasible. Also, $b^T y^1 = b^T y^2 = ... = b^T y^{L+1}$ because the construction preserves objective. Finally, $P_{B_l}(K(y^l)) = K(y^{l+1})$ for all $l \in [L]$, therefore $P_{B_L}(K(y^L)) = K(y^{L+1}) = K$ and $P_{\mathcal{B}}(K) = K$, so $P_{\mathcal{B}}(K(y^{L+1}) = K(y^{L+1})$. By Corollary 2, y^{L+1} is an ILM w.r.t. $\mathcal{B}$.

By Remark 1, the sequence $y^1, ..., y^{L+1}$ can be obtained by updating the corresponding blocks into the relative interior of optimizers. Because the objective did not improve during these updates and y^{L+1} is ILM, it follows from Theorem 1 (statements A,C,D) that y is a pre-ILM. $\square$

Proposition 10. *If y is a pre-ILM of dual* (1) *w.r.t. $\mathcal{B}$, then $P_{\mathcal{B}}(K(y)) \neq \perp$.*

Proof. Proof by contradiction. Suppose y is pre-ILM and $P_{\mathcal{B}}(K(y)) = \perp$, then there must exist a finite sequence $(B_l)_{l=1}^{L}$ for $B_l \in \mathcal{B}, l \in [L]$ such that

$$P_{B_L}(P_{B_{L-1}}(P_{B_{L-2}}(\cdots P_{B_2}(P_{B_1}(K(y)))\cdots))) = \perp \tag{24}$$

which consists of the used blocks B in the propagation algorithm.

As discussed in Remark 1, we can imitate this propagation by creating a sequence of dual feasible points $y^1, y^2, \cdots, y^L$ where $y^1 = y$ and y^{l+1} is created from y^l by changing block of variables B_l to be in the relative interior of optimizers. This is given by construction in the proof of Proposition 8 and it holds that $P_{B_l}(K(y^l)) = K(y^{l+1})$ for all $l \in [L-1]$. Since $P_{B_L}(K(y^L)) = \perp$, y^L is not a local minimum by Proposition 6. Therefore, updating the block of variables B_L in y^L by (4) (or even (3)) to obtain a point y^{L+1} improves objective.

Thus, we applied BCD with relative interior rule to obtain the sequence $(y^l)_{l=1}^{L+1}$ and the objective improved. This is contradictory with Theorem 1 (statement C) which states that block updates that choose any optimizer (even without relative interior rule) cannot improve the objective from a pre-ILM. $\square$

*Proof (**Theorem 2**).* For the first part, point y is an LM of dual (1) w.r.t. $\mathcal{B}$ by its definition if y_B satisfies (3) for all $B \in \mathcal{B}$. By Proposition 6, this is equivalent to $P_B(K(y)) \neq \bot\ \forall B \in \mathcal{B}$. The second part is given in Corollary 2 and the third part follows from Proposition 9 and Proposition 10. $\square$

B Constructing an Improving Feasible Direction

As discussed in §3, if (6) is infeasible, there exists an improving feasible direction (7), we are going to describe how to obtain it based on the propagation algorithm defined in §3.1. We remark that conditions (7) define a whole convex cone of improving directions and our algorithm finds one of them based on the specific implementation of the construction.

Let us have a set of blocks $\mathcal{B} \subseteq 2^{[m]}$ and a dual feasible point y such that $P_{\mathcal{B}}(K(y)) = \bot$, which implies infeasibility of (6). Consider sequences $(B_l)_{l=1}^L$ and $(K_l)_{l=1}^L$ where $K_1 \supset K_2 \supset \cdots \supset K_L$, $K_1 = K(y)$, $K_{l+1} = P_{B_l}(K_l)$ for every $l \in [L-1]$, and $P_{B_L}(K_L) = \bot$. To construct $\bar{y}$, we use the primal-dual pair

$$
\begin{aligned}
&\max 0 && \min b^T \bar{y}^l && && \text{(25a)}\\
&A^j x = b_j && \bar{y}^l_j \in \mathbb{R} && \forall j \in B_l && \text{(25b)}\\
&x_i = 0 && - && \forall i \in [n] - K_l && \text{(25c)}\\
&x_i \geq 0 && A_i^T \bar{y}^l \geq 0 && \forall i \in K_l && \text{(25d)}\\
&- && \bar{y}^l_j = 0 && \forall j \in [m] - B_l. && \text{(25e)}
\end{aligned}
$$

and proceed as follows:

1. Initialize $\bar{y} \leftarrow \bar{y}^L$ where $\bar{y}^L$ is any feasible dual solution of (25) for $l = L$ with[7] $b^T \bar{y}^L < 0$.
2. For all $l \in \{L-1, L-2, ..., 2, 1\}$ in descending order:
   (a) If $A_i^T \bar{y} \geq 0$ for all $i \in K_l - K_{l+1}$, continue with $l \leftarrow l - 1$.
   (b) Else, find $\bar{y}^l$ from the relative interior of optimizers of dual (25) for current l, update $\bar{y} \leftarrow \bar{y} + \delta_l \bar{y}^l$ where $\delta_l = \max\limits_{\substack{i \in K_l - K_{l+1} \\ A_i^T \bar{y} < 0}} -\frac{A_i^T \bar{y}}{A_i^T \bar{y}^l}$, and set $l \leftarrow l - 1$.
3. Return $\bar{y}$ as improving feasible direction satisfying (7).

Due to lack of space, we omit the proof of this procedure. We will only state that it is based on induction, i.e., after some index $l \in [L]$ is processed, it holds that $A_i^T \bar{y} \geq 0$ for all $i \in K_l$ and $b^T \bar{y} = b^T \bar{y}^L < 0$ is maintained during the whole algorithm. Thus, eventually $A_i^T \bar{y} \geq 0$ holds for all $i \in K_1 = K(y)$.

After $\bar{y}$ is calculated, we can find a step size $\epsilon > 0$ and perform update of y as discussed in §3. Even though this approach may seem complicated, it is easy to see that in cases when the blocks B are small, the problem (25) is also small and could even be solvable in closed-form for some special cases.

[7] Such $\bar{y}^L$ exists because primal (25) is infeasible for $l = L$ due to $P_{B_L}(K_L) = \bot$ and the dual (25) is therefore unbounded since the dual always has a feasible solution.

References

1. Bessiere, C.: Constraint propagation. In: Handbook of Constraint Programming, chap. 3. Elsevier (2006)
2. Cooper, M.C., de Givry, S., Sanchez, M., Schiex, T., Zytnicki, M., Werner, T.: Soft arc consistency revisited. Artificial Intelligence **174**(7-8), 449–478 (2010)
3. Dlask, T.: Minimizing convex piecewise-affine functions by local consistency techniques. Master's Thesis (2018)
4. Dlask, T.: Unit propagation by means of coordinate-wise minimization. In: International Conference on Machine Learning, Optimization, and Data Science. Springer (2020)
5. Dlask, T., Werner, T.: Bounding linear programs by constraint propagation: Application to Max-SAT. In: International Conference on Principles and Practice of Constraint Programming. Springer (2020)
6. Dlask, T., Werner, T.: A class of linear programs solvable by coordinate-wise minimization. In: Kotsireas, I.S., Pardalos, P.M. (eds.) Learning and Intelligent Optimization. pp. 52–67. Springer (2020)
7. Globerson, A., Jaakkola, T.S.: Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In: Advances in Neural Information Processing Systems. pp. 553–560 (2008)
8. Kappes, J.H., Andres, B., Hamprecht, F.A., Schnörr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B.X., Kröger, T., Lellmann, J., Komodakis, N., Savchynskyy, B., Rother, C.: A comparative study of modern inference techniques for structured discrete energy minimization problems. Intl. J. of Computer Vision **115**(2), 155–184 (2015)
9. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(10), 1568–1583 (2006)
10. Koval, V.K., Schlesinger, M.I.: Dvumernoe programmirovanie v zadachakh analiza izobrazheniy (Two-dimensional programming in image analysis problems). Automatics and Telemechanics **8**, 149–168 (1976), in Russian.
11. Kovalevsky, V., Koval, V.: A diffusion algorithm for decreasing energy of max-sum labeling problem. Glushkov Institute of Cybernetics, Kiev, USSR (1975), unpublished.
12. Lemaréchal, C., Hiriart-Urruty, J.: Fundamentals of Convex Analysis. Springer Grundlehren Text Editions, Springer Verlag, New York (2004)
13. Matoušek, J., Gärtner, B.: Understanding and using linear programming (universitext). Springer-Verlag (2006)
14. Papadimitriou, C.H., Steiglitz, K.: Combinatorial optimization: algorithms and complexity. Courier Corporation (1998)
15. Savchynskyy, B.: Discrete graphical models – an optimization perspective. Foundations and Trends in Computer Graphics and Vision **11**(3-4), 160–429 (2019)
16. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. J. Optim. Theory Appl. **109**(3), 475–494 (Jun 2001)
17. Živný, S.: The Complexity of Valued Constraint Satisfaction Problems. Cognitive Technologies, Springer (2012)
18. Wainwright, M.J., Jordan, M.I.: Graphical models, exponential families, and variational inference. Foundations and Trends in Machine Learning **1**(1-2), 1–305 (2008)
19. Werner, T.: A linear programming approach to max-sum problem: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(7), 1165–1179 (July 2007)

20. Werner, T.: On coordinate minimization of piecewise-affine functions. Tech. Rep. CTU-CMP-2017-05, Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague (September 2017)
21. Werner, T., Průša, D., Dlask, T.: Relative interior rule in block-coordinate descent. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7559–7567 (2020)
22. Werner, T., Průša, D.: Relative interior rule in block-coordinate minimization. ArXiv.org (2019)
23. Zhang, S.: On the strictly complementary slackness relation in linear programming. In: Advances in Optimization and Approximation, pp. 347–361. Springer (1994)