# A Class of Linear Programs Solvable by Coordinate-wise Minimization

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

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

**Abstract.** Coordinate-wise minimization is a simple popular method for large-scale optimization. Unfortunately, for general (non-differentiable and/or constrained) convex problems it may not find global minima. We present a class of linear programs that coordinate-wise minimization solves exactly. We show that dual LP relaxations of several well-known combinatorial optimization problems are in this class and the method finds a global minimum with sufficient accuracy in reasonable runtimes. Moreover, for extensions of these problems that no longer are in this class the method yields reasonably good suboptima. Though the presented LP relaxations can be solved by more efficient methods (such as max-flow), our results are theoretically non-trivial and can lead to new large-scale optimization algorithms in the future.

**Keywords:** Coordinate-wise minimization · Linear programming · LP relaxation

## 1 Introduction

*Coordinate-wise minimization*, or *coordinate descent*, is an iterative optimization method, which in every iteration optimizes only over a single chosen variable while keeping the remaining variables fixed. Due its simiplicity, this method is popular among practitioners in large-scale optimization in areas such as machine learning or computer vision, see e.g. [32]. A natural extension of the method is *block-coordinate minimization*, where every iteration minimizes the objective over a block of variables. In this paper, we focus on coordinate minimization with exact updates, where in each iteration a global minimum over the chosen variable is found, applied to convex optimization problems.

For general convex optimization problems, the method need not converge and/or its fixed points need not be global minima. A simple example is the unconstrained minimization of the function $f(x,y) = \max\{x - 2y, y - 2x\}$, which is unbounded but any point with $x = y$ is a coordinate-wise local minimum. Despite this drawback, (block-)coordinate minimization can be very successful for some large-scale convex non-differentiable problems. The prominent example is the class of *convergent message passing* methods for solving dual linear programming (LP) relaxation of maximum a posteriori (MAP) inference in graphical models, which can be seen as various forms of (block-)coordinate

descent applied to various forms of the dual. In the typical case, the dual LP relaxation boils down to the unconstrained minimization of a convex piece-wise affine (hence non-differentiable) function. These methods include max-sum diffusion [21, 26, 29], TRW-S [18], MPLP [12], and SRMP [19]. They do not guarantee global optimality but for large sparse instances from computer vision the achieved coordinate-wise local optima are very good and TRW-S is significantly faster than competing methods [27, 16], including popular first-order primal-dual methods such as ADMM [5] or [8].

This is a motivation to look for other classes of convex optimization problems for which (block-)coordinate descent would work well or, alternatively, to extend convergent message passing methods to a wider class of convex problems than the dual LP relaxation of MAP inference. A step in this direction is the work [31], where it was observed that if the minimizer of the problem over the current variable block is not unique, one should choose a minimizer that lies in the *relative interior* of the set of block-optimizers. It is shown that any update satisfying this rule is, in a precise sense, not worse than any other exact update. Message-passing methods such as max-sum diffusion and TRW-S satisfy this rule. If max-sum diffusion is modified to violate the relative interior rule, it can quickly get stuck in a very poor coordinate-wise local minimum.

To be precise, suppose we minimize a convex function $f \colon X \to \mathbb{R}$ on a closed convex set $X \subseteq \mathbb{R}^n$. We assume that $f$ is bounded from below on $X$. For brevity of formulation, we rephrase this as the minimization of the extended-valued function $\bar{f} \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ such that $\bar{f}(x) = f(x)$ for $x \in X$ and $\bar{f}(x) = \infty$ for $x \notin X$. One iteration of coordinate minimization with the relative interior rule [31] chooses a variable index $i \in [n] = \{1, \ldots, n\}$ and replaces an estimate $x^k = (x_1^k, \ldots, x_n^k) \in X$ with a new estimate $x^{k+1} = (x_1^{k+1}, \ldots, x_n^{k+1}) \in X$ such that[1]

$$x_i^{k+1} \in \operatorname{ri} \operatorname*{argmin}_{y \in \mathbb{R}} \bar{f}(x_1^k, \ldots, x_{i-1}^k, y, x_{i+1}^k, \ldots, x_n^k),$$
$$x_j^{k+1} = x_j^k \quad \forall j \neq i,$$

where $\operatorname{ri} Y$ denotes the relative interior of a convex set $Y$. As this is a univariate convex problem, the set $Y = \operatorname{argmin}_{y \in \mathbb{R}} \bar{f}(x_1^k, \ldots, x_{i-1}^k, y, x_{i+1}^k, \ldots, x_n^k)$ is either a singleton or an interval. In the latter case, the relative interior rule requires that we choose $x_i^{k+1}$ from the interior of this interval. A point $x = (x_1, \ldots, x_n) \in X$ that satisfies

$$x_i \in \operatorname{ri} \operatorname*{argmin}_{y \in \mathbb{R}} \bar{f}(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n)$$

for all $i \in [n]$ is called a (coordinate-wise) *interior local minimum* of function $f$ on set $X$.

Some classes of convex problems are solved by coordinate-wise minimization exactly. E.g., for unconstrained minimization of a differentiable convex function,

---

[1] In [31], the iteration is formulated in a more abstract (coordinate-free) notation. Since we focus only on coordinate-wise minimization here, we use a more concrete notation.

it is easy to see that any fixed point of the method is a global minimum; moreover, it has been proved that if the function has unique univariate minima, then any limit point is a global minimum [4, §2.7]. The same properties hold for convex functions whose non-differentiable part is separable [28]. Note that these classical results need not assume the relative interior rule [31].

Therefore, it is natural to ask if the relative interior rule can widen the class of convex optimization problems that are exactly solved by coordinate-wise minimization. Leaving convergence aside[2], more precisely we can ask for which problems interior local minima are global minima. A succinct characterization of this class is currently out of reach. Two subclasses of this class are known [18, 26, 29]: the dual LP relaxation of MAP inference with pairwise potential functions and two labels, or with submodular potential functions.

In this paper, we restrict ourselves to linear programs (where $f$ is linear and $X$ is a convex polyhedron) and present a new class of linear programs with this property. We show that dual LP relaxations of a number of combinatorial optimization problems belong to this class and coordinate-wise minimization converges in reasonable time on large practical instances. Unfortunately, the practical impact of this result is limited because there exist more efficient algorithms for solving these LP relaxations, such as reduction to max-flow. It is open whether there exist some useful classes of convex problems that are exactly solvable by (block-)coordinate descent but not solvable by more efficient methods. There is a possibility that our result and the proof technique will pave the way to such results.

## 2 Reformulations of Problems

Before presenting our main result, we make an important remark: while a convex optimization problem can be reformulated in many ways to an 'equivalent' problem which has the same global minima, not all of these transformations are equivalent with respect to coordinate-wise minimization, in particular, not all preserve interior local minima.

*Example 1.* One example is dualization. If coordinate-wise minimization achieves good local (or even global) minima on a convex problem, it can get stuck in very poor local minima if applied to its dual. Indeed, trying to apply (block-) co-ordinate minimization to the *primal* LP relaxation of MAP inference (linear optimization over the local marginal polytope) has been futile so far.

*Example 2.* Consider the linear program $\min\{x_1 + x_2 \mid x_1, x_2 \geq 0\}$, which has one interior local minimum with respect to individual coordinates that also corresponds to the unique global optimum. But if one adds a redundant constraint, namely $x_1 = x_2$, then any feasible point will become an interior local minimum

---

[2] We do not discuss convergence in this paper and assume that the method converges to an interior local minimum. This is supported by experiments, e.g., max-sum diffusion and TRW-S have this property. More on convergence can be found in [31].

w.r.t. individual coordinates, because the redundant constraint blocks changing the variable $x_i$ without changing $x_{3-i}$ for both $i \in \{1, 2\}$.

*Example 3.* Consider the linear program

$$\min \sum_{j=1}^{m} z_j \tag{1a}$$

$$z_j \geq a_{ij}^T x + b_{ij} \qquad \forall i \in [n], j \in [m] \tag{1b}$$

$$z \in \mathbb{R}^m, x \in \mathbb{R}^p \tag{1c}$$

which can be also formulated as

$$\min \sum_{j=1}^{m} \max_{i=1}^{n} (a_{ij}^T x + b_{ij}) \tag{2a}$$

$$x \in \mathbb{R}^p. \tag{2b}$$

Optimizing over the individual variables by coordinate-wise minimization in (1) does not yield the same interior local optima as in (2). For instance, assume that $m = 3$, $n = p = 1$ and the problem (2) is given as

$$\min \left( \max\{x, 0\} + \max\{-x, -1\} + \max\{-x, -2\} \right), \tag{3}$$

where $x \in \mathbb{R}$. Then, when optimizing directly in form (3), one can see that all the interior local optima are global optimizers.

However, when one introduces the variables $z \in \mathbb{R}^3$ and applies coordinate-wise minimization on the corresponding problem (1), then there are interior local optima that are not global optimizers, for example $x = z_1 = z_2 = z_3 = 0$, which is an interior local optimum, but is not a global optimum.

On the other hand, optimizing over blocks of variables $\{z_1, \ldots, z_m, x_i\}$ for each $i \in [p]$ in case (1) is equivalent to optimization over individual $x_i$ in formulation (2).

## 3   Main Result

The optimization problem with which we are going to deal is in its most general form defined as

$$\min \left( \sum_{i=1}^{m} \max\{w_i - \varphi_i, 0\} + a^T \varphi + b^T \lambda + \sum_{j=1}^{p} \max\{v_j + A_{:j}^T \varphi + B_{:j}^T \lambda, 0\} \right) \tag{4a}$$

$$\underline{\varphi}_i \leq \varphi_i \leq \overline{\varphi}_i \quad \forall i \in [m] \tag{4b}$$

$$\underline{\lambda}_i \leq \lambda_i \leq \overline{\lambda}_i \quad \forall i \in [n], \tag{4c}$$

where $A \in \mathbb{R}^{m \times p}, B \in \mathbb{R}^{n \times p}, a \in \mathbb{R}^m, b \in \mathbb{R}^n, w \in \mathbb{R}^m, v \in \mathbb{R}^p, \underline{\varphi} \in (\mathbb{R} \cup \{-\infty\})^m, \overline{\varphi} \in (\mathbb{R} \cup \{\infty\})^m, \underline{\lambda} \in (\mathbb{R} \cup \{-\infty\})^n, \overline{\lambda} \in (\mathbb{R} \cup \{\infty\})^n$ (assuming $\underline{\varphi} < \overline{\varphi}$

and $\underline{\lambda} < \overline{\lambda}$). We optimize over variables $\varphi \in \mathbb{R}^m$ and $\lambda \in \mathbb{R}^n$. $A_{:j}$ and $A_{i:}$ denotes the $j$-th column and $i$-th row of $A$, respectively.

Applying coordinate-wise minimization with relative-interior rule on the problem (4) corresponds to cyclic updates of variables, where each update corresponds to finding the region of optima of a convex piecewise-affine function of one variable on an interval. If the set of optimizers is a singleton, then the update is straightforward. If the set of optimizers is a bounded interval $[a, b]$, the variable is assigned the middle value from this interval, i.e. $(a + b)/2$. If the set of optima is unbounded, i.e. $[a, \infty)$, then we set the variable to the value $a + \Delta$, where $\Delta > 0$ is a fixed constant. In case of $(-\infty, a]$, the variable is updated to $a - \Delta$. The details for the update in this setting are in Appendix A in [10].

**Theorem 1.** *Any interior local optimum of* (4) *w.r.t. individual coordinates is its global optimum if*

- *matrices $A, B$ contain only values from the set $\{-1, 0, 1\}$ and contain at most two non-zero elements per row*
- *vector $a$ contains only elements from the set $(-\infty, -2] \cup \{-1, 0, 1, 2\} \cup [3, \infty)$*
- *vector $b$ contains only elements from the set $(-\infty, -2] \cup \{-1, 0, 1\} \cup [2, \infty)$.*

In order to prove Theorem 1, we formulate problem (4) as a linear program by introducing additional variables $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{R}^p$ and construct its dual. The proof of optimality is then obtained (see Theorem 2) by constructing a dual feasible solution that satisfies complementary slackness.

The primal linear program (with corresponding dual variables and constraints on the same lines) reads

$$\min \sum_{i \in [m]} \alpha_i + \sum_{i \in [p]} \beta_i + a^T \varphi + b^T \lambda \qquad \max f(z, y, s, r, q, x) \tag{5a}$$

$$\beta_j - A_{:j}^T \varphi - B_{:j}^T \lambda \geq v_j \qquad\qquad\qquad x_j \geq 0 \quad \forall j \in [p] \quad \text{(5b)}$$

$$\alpha_i + \varphi_i \geq w_i \qquad\qquad\qquad s_i \geq 0 \quad \forall i \in [m] \quad \text{(5c)}$$

$$\varphi_i \geq \underline{\varphi}_i \qquad\qquad\qquad y_i \geq 0 \quad \forall i \in [m] \quad \text{(5d)}$$

$$\varphi_i \leq \overline{\varphi}_i \qquad\qquad\qquad z_i \leq 0 \quad \forall i \in [m] \quad \text{(5e)}$$

$$\lambda_i \geq \underline{\lambda}_i \qquad\qquad\qquad q_i \geq 0 \quad \forall i \in [n] \quad \text{(5f)}$$

$$\lambda_i \leq \overline{\lambda}_i \qquad\qquad\qquad r_i \leq 0 \quad \forall i \in [n] \quad \text{(5g)}$$

$$\varphi_i \in \mathbb{R} \qquad\qquad s_i + z_i + y_i - A_{i:}^T x = a_i \quad \forall i \in [m] \quad \text{(5h)}$$

$$\lambda_i \in \mathbb{R} \qquad\qquad r_i + q_i - B_{i:}^T x = b_i \quad \forall i \in [n] \quad \text{(5i)}$$

$$\beta_j \geq 0 \qquad\qquad\qquad x_j \leq 1 \quad \forall j \in [p] \quad \text{(5j)}$$

$$\alpha_i \geq 0 \qquad\qquad\qquad s_i \leq 1 \quad \forall i \in [m], \quad \text{(5k)}$$

where the dual criterion is

$$f(z, y, s, r, q, x) = \overline{\varphi}^T z + \underline{\varphi}^T y + w^T s + \overline{\lambda}^T r + \underline{\lambda}^T q + v^T x \tag{6}$$

and clearly, at optimum of the primal, we have

$$\alpha_i = \max\{w_i - \varphi_i, 0\} \qquad\qquad \forall i \in [m] \qquad (7a)$$

$$\beta_j = \max\{v_j + A_{:j}^T\varphi + B_{:j}^T\lambda, 0\} \qquad\qquad \forall j \in [p]. \qquad (7b)$$

The variables $\alpha, \beta$ were eliminated from the primal formulation (5) to obtain (4) due to similar reasoning as in Example 3. We also remark that setting $\overline{\varphi}_i = \infty$ (resp. $\underline{\varphi}_i = -\infty$, $\overline{\lambda}_i = \infty$, $\underline{\lambda}_i = -\infty$) results in $z_i = 0$ (resp. $y_i = 0$, $r_i = 0$, $q_i = 0$).

   Even though the primal-dual pair (5) might seem overcomplicated, such general description is in fact necessary because as described in §2, equivalent reformulations may not preserve the structure of interior local minima and we would like to describe as general class, where optimality is guaranteed, as possible.

*Example 4.* To give the reader better insight into the problems (5), we present a simplification based on omitting the matrix $A$ (i.e. $m = 0$) and setting $\underline{\lambda} = 0$, $\overline{\lambda} = \infty$, which results in $r_i = 0$ and variables $q_i$ become slack variables in (5i). The primal-dual pair in this case then simplifies to

$$\min \sum_{i \in [p]} \beta_i + b^T\lambda \qquad\qquad \max v^T x \qquad\qquad (8a)$$

$$\beta_j - B_{:j}^T\lambda \geq v_j \qquad\qquad x_j \geq 0 \qquad\qquad \forall j \in [p] \qquad (8b)$$

$$\beta_j \geq 0 \qquad\qquad x_j \leq 1 \qquad\qquad \forall j \in [p] \qquad (8c)$$

$$\lambda_i \geq 0 \qquad\qquad -B_{i:}^T x \leq b_i \qquad\qquad \forall i \in [n]. \qquad (8d)$$

**Theorem 2.** *For a problem* (4) *satisfying conditions of Theorem 1 and a given interior local minimum* $(\varphi, \lambda)$, *the values*[3]

$$x_j = \begin{cases} 0 & \text{if } A_{:j}^T\varphi + B_{:j}^T\lambda + v_j < 0 \\ \frac{1}{2} & \text{if } A_{:j}^T\varphi + B_{:j}^T\lambda + v_j = 0 \\ 1 & \text{if } A_{:j}^T\varphi + B_{:j}^T\lambda + v_j > 0 \end{cases} \qquad s_i = \begin{cases} 1 & \text{if } w_i > \varphi_i \\ 0 & \text{if } w_i < \varphi_i \\ h_{[0,1]}(a_i + A_{i:}^T x) & \text{if } w_i = \varphi_i \end{cases}$$

$$r_i = \begin{cases} 0 & \text{if } \lambda_i < \overline{\lambda}_i \\ h_{\mathbb{R}_0^-}(b_i + B_{i:}^T x) & \text{if } \lambda_i = \overline{\lambda}_c \end{cases} \qquad z_i = \begin{cases} 0 & \text{if } \varphi_i < \overline{\varphi}_i \\ h_{\mathbb{R}_0^-}(a_i + A_{i:}^T x - s_i) & \text{if } \varphi_i = \overline{\varphi}_i \end{cases}$$

$$q_i = \begin{cases} 0 & \text{if } \lambda_i > \underline{\lambda}_i \\ h_{\mathbb{R}_0^+}(b_i + B_{i:}^T x) & \text{if } \lambda_i = \underline{\lambda}_i \end{cases} \qquad y_i = \begin{cases} 0 & \text{if } \varphi_i > \underline{\varphi}_i \\ h_{\mathbb{R}_0^+}(a_i + A_{i:}^T x - s_i) & \text{if } \varphi_i = \underline{\varphi}_i \end{cases}$$

*are feasible for the dual* (5) *and satisfy complementary slackness with primal* (5), *where the remaining variables of the primal are given by* (7).

   It can be immediately seen that all the constraints of dual (5) are satisfied except for (5h) and (5i), which require a more involved analysis. The complete proof of Theorem 2 is technical (based on verifying many different cases) and given in Appendix B in [10].

---

[3] We define $h_{[x,y]}(z) = \min\{y, \max\{z, x\}\}$ to be the projection of $z \in \mathbb{R}$ onto the interval $[x, y] \subseteq \mathbb{R}$. The projection onto unbounded intervals $(-\infty, 0]$ and $[0, \infty)$ is defined similarly and is denoted by $h_{\mathbb{R}_0^-}$ and $h_{\mathbb{R}_0^+}$ for brevity.

## 4    Applications

Here we show that several LP relaxations of combinatorial problems correspond to the form (4) or to the dual (5) and discuss which additional constraints correspond to the assumptions of Theorem 1.

### 4.1    Weighted Partial Max-SAT

In weighted partial Max-SAT, one is given two sets of clauses, soft and hard. Each soft clause is assigned a positive weight. The task is to find values of binary variables $x_i \in \{0, 1\}$, $i \in [p]$ such that all the hard clauses are satisfied and the sum of weights of the satisfied soft clauses is maximized.

We organize the $m$ soft clauses into a matrix $S \in \{-1, 0, 1\}^{m \times p}$ defined as

$$S_{ci} = \begin{cases} 1 & \text{if literal } x_i \text{ is present in soft clause } c \\ -1 & \text{if literal } \neg x_i \text{ is present in soft clause } c \ , \\ 0 & \text{otherwise} \end{cases}$$

In addition, we denote $n_c^S = \sum_i \llbracket S_{ci} < 0 \rrbracket$ to be the number of negated variables in clause $c$. These numbers are stacked in a vector $n^S \in \mathbb{Z}^m$. The $h$ hard clauses are organized in a matrix $H \in \{-1, 0, 1\}^{h \times p}$ and a vector $n^H \in \mathbb{Z}^h$ in the same manner.

The LP relaxation of this problem reads

$$\max \sum_{c \in [m]} w_c s_c \tag{10a}$$

$$s_c \leq S_{c:}^T x + n_c^S \qquad \forall c \in [m] \tag{10b}$$

$$H_{c:}^T x + n_c^H \geq 1 \qquad \forall c \in [h] \tag{10c}$$

$$x_i \in [0, 1] \qquad \forall i \in [p] \tag{10d}$$

$$s_c \in [0, 1] \qquad \forall c \in [m], \tag{10e}$$

where $w_c \in \mathbb{R}_0^+$ are the weights of the soft clauses $c \in [m]$. This is a sub-class of the dual (5), where $A = S$, $B = -H$, $a = n^S$, $b = 1 - n^H$, $\varphi = 0$ ($y \geq 0$ are therefore slack variables for the dual constraint (5h) that correspond to (10b)), $\overline{\varphi} = \infty$ (therefore $z = 0$), $\underline{\lambda} = -\infty$ (therefore $q = 0$), $\overline{\lambda} = 0$ ($r \leq 0$ are slack variables for the dual constraint (5i) that correspond to (10c)), $v = 0$.

Formulation (10) satisfies the conditions of Theorem 1 if each of the clauses has length at most 2. In other words, optimality is guaranteed for weighted partial Max-2SAT.

Also notice that if we omitted the soft clauses (10b) and instead set $v = -1$, we would obtain an instance of Min-Ones SAT, which could be generalized to weighted Min-Ones SAT. This relaxation would still satisfy the requirements of Theorem 1 if all the present hard clauses have length at most 2.

**Results** We tested the method on 800 smallest[4] instances that appeared in Max-SAT Evaluations [2] in years 2017 [1] and 2018 [3]. The results on the instances are divided into groups in Table 1 based on the minimal and maximal length of present clauses. We have also tested this approach on 60 instances of weighted Max-2SAT from Ke Xu [33]. The highest number of logical variables in an instance was 19034 and the highest overall number of clauses in an instance was 31450. It was important to separate the instances without unit clauses (i.e. clauses of length 1), because in such cases the LP relaxation (10) has a trivial optimal solution with $x_i = \frac{1}{2}$ for all $i \in V$.

Coordinate-wise minimization was stopped when the criterion did not improve by at least $\epsilon = 10^{-7}$ after a whole cycle of updates for all variables. We report the quality of the solution as the median and mean relative difference between the optimal criterion and the criterion reached by coordinate-wise minimization before termination.

Table 1 reports not only instances of weighted partial Max-2SAT but also instances with longer clauses, where optimality is no longer guaranteed. Nevertheless, the relative differences on instances with longer clauses still seem not too large and could be usable as bounds in a branch-and-bound scheme.

| Instance Group Specification | | | Results | |
|---|---|---|---|---|
| Min CL | Max CL | #inst. | Mean RD | Median RD |
| $\geq 2$ | any | 91 | 0 | 0 |
| 1 | 2 | 123 | $1.44 \cdot 10^{-9}$ | $1.09 \cdot 10^{-11}$ |
| 1 | 3 | 99 | $6.98 \cdot 10^{-3}$ | $1.90 \cdot 10^{-7}$ |
| 1 | $\geq 4$ | 487 | $1.26 \cdot 10^{-2}$ | $2.97 \cdot 10^{-3}$ |
| 1 | 2 | 60 | $1.59 \cdot 10^{-9}$ | $5.34 \cdot 10^{-10}$ |

**Table 1.** Experimental comparison of coordinate-wise minimization and exact solutions for LP relaxation on instances from [2] (first 4 rows) and [33] (last row).

### 4.2   Weighted Vertex Cover

Dual (8) also subsumes[5] the LP relaxation of weighted vertex cover, which reads

$$\min \left\{ \sum_{i \in V} v_i x_i \mid x_i + x_j \geq 1 \; \forall \{i, j\} \in E, x_i \in [0, 1] \; \forall i \in V \right\} \qquad (11)$$

where $V$ is the set of nodes and $E$ is the set of edges of an undirected graph. This problem also satisfies the conditions of Theorem 1 and therefore the corresponding primal (4) will have no non-optimal interior local minima.

---

[4] Smallest in the sense of the file size. All instances could not have been evaluated due to their size and lengthy evaluation.

[5] It is only necessary to transform minimization to maximization of negated objective in (11).

On the other hand, notice that formulation (11), which corresponds to dual (5) can have non-optimal interior local minima even with respect to all subsets of variables of size $|V| - 1$, an example is given in Appendix C in [10].

We reported the experiments on weighted vertex cover in [30] where the optimality was not proven yet. In addition, the update designed in [30] *ad hoc* becomes just a special case of our general update here.

### 4.3 Minimum $st$-Cut, Maximum Flow

Recall from [11] the usual formulation of max-flow problem between nodes $s \in V$ and $t \in V$ on a directed graph with vertex set $V$, edge set $E$ and positive edge weights $w_{ij} \in \mathbb{R}_0^+$ for each $(i, j) \in E$, which reads

$$\max \sum_{(s,i) \in E} f_{si} \tag{12a}$$

$$0 \le f_{ij} \le w_{ij} \qquad \forall (i,j) \in E \tag{12b}$$

$$\sum_{(u,i) \in E} f_{ui} - \sum_{(j,u) \in E} f_{ju} = 0 \qquad \forall u \in V - \{s, t\}. \tag{12c}$$

Assume that there is no edge $(s, t)$, there are no ingoing edges to $s$ and no outgoing edges from $t$, then any feasible value of $f$ in (12) is an interior local optimum w.r.t. individual coordinates by the same reasoning as in Example 2 due to the flow conservation constraint (12c), which limits each individual variable to a single value. We are going to propose a formulation which has no non-globally optimal interior local optima.

The dual problem to (12) is the minimum $st$-cut problem, which can be formulated as

$$\max \sum_{(i,j) \in E} w_{ij} y_{ij} \tag{13a}$$

$$y_{ij} \le 1 - x_i + x_j \qquad \forall (i,j) \in E, i \ne s, j \ne t \tag{13b}$$

$$y_{sj} \le x_j \qquad \forall (s,j) \in E \tag{13c}$$

$$y_{it} \le 1 - x_i \qquad \forall (i,t) \in E \tag{13d}$$

$$y_{ij} \in [0, 1] \qquad \forall (i,j) \in E, \tag{13e}$$

$$x_i \in [0, 1] \qquad \forall i \in V - \{s, t\}, \tag{13f}$$

where $y_{ij} = 0$ if edge $(i, j)$ is in the cut and $y_{ij} = 1$ if edge $(i, j)$ is not in the cut. The cut should separate $s$ and $t$, so the set of nodes connected to $s$ after the cut will be denoted by $S$ and $T = V - S$ is the set of nodes connected to $t$. Using this notation, $x_i = [\![ i \in S ]\!]$. Formulation (13) is different from the usual formulation by replacing the variables $y_{ij}$ by $1 - y_{ij}$, therefore we also maximize the weight of the not cut edges instead of minimizing the weight of the cut edges, therefore if the optimal value of (13) is $O$, then the value of the minimum $st$-cut equals $\sum_{(i,j) \in E} w_{ij} - O$.

Formulation (13) is subsumed by the dual (5) by setting $\underline{\varphi} = 0$, $\overline{\varphi} = \infty$ and omitting the $B$ matrix. Also notice that each $y_{ij}$ variable occurs in at most one constraint. The problem (13) therefore satisfies the conditions of Theorem 1 and the corresponding primal (4) is a formulation of the maximum flow problem, in which one can search for the maximum flow by coordinate-wise minimization. The corresponding formulation (4) reads

$$\min\Big( \sum_{(i,j)\in E} \max\{w_{ij} - \varphi_{ij}, 0\} + \sum_{(i,j)\in E, i\neq s} \varphi_{ij} +$$
$$+ \sum_{i\in V-\{s,t\}} \max\Big\{ \sum_{(j,i)\in E} \varphi_{ji} - \sum_{(i,j)\in E} \varphi_{ij}, 0 \Big\}\Big) \tag{14a}$$
$$\varphi_{ij} \geq 0 \ \ \forall (i,j) \in E. \tag{14b}$$

**Results** We have tested our formulation for coordinate-wise minimization on max-flow instances[6] from computer vision. We report the same statistics as with Max-SAT in Table 2, the instances corresponded to stereo problems, multiview reconstruction instances and shape fitting problems.

For multiview reconstruction and shape fitting, we were able to run our algorithm only on small instances, which have approximately between $8 \cdot 10^5$ and $1.2 \cdot 10^6$ nodes and between $5 \cdot 10^6$ and $6 \cdot 10^6$ edges. On these instances, the algorithm terminated with the reported precision in 13 to 34 minutes on a laptop.

| Instance Group or Instance | | Results | |
|---|---|---|---|
| Name | #inst | Mean RD | Median RD |
| BVZ-tsukuba [7] | 16 | $6.03 \cdot 10^{-10}$ | $1.17 \cdot 10^{-11}$ |
| BVZ-sawtooth [25] [7] | 20 | $9.83 \cdot 10^{-11}$ | $6.11 \cdot 10^{-12}$ |
| BVZ-venus [25] [7] | 22 | $3.40 \cdot 10^{-11}$ | $2.11 \cdot 10^{-12}$ |
| KZ2-tsukuba [20] | 16 | $2.69 \cdot 10^{-10}$ | $1.77 \cdot 10^{-10}$ |
| KZ2-sawtooth [25] [20] | 20 | $4.08 \cdot 10^{-9}$ | $1.56 \cdot 10^{-10}$ |
| KZ2-venus [25] [20] | 22 | $5.21 \cdot 10^{-9}$ | $1.74 \cdot 10^{-10}$ |
| BL06-camel-sml [23] | 1 | $1.21 \cdot 10^{-11}$ | |
| BL06-gargoyle-sml [6] | 1 | $6.29 \cdot 10^{-12}$ | |
| LB07-bunny-sml [22] | 1 | $1.33 \cdot 10^{-10}$ | |

**Table 2.** Experimental comparison of coordinate-wise minimization on max-flow instances, the references are the original sources of the data and/or to the authors that reformulated these problems as maximum flow. The first 6 rows correspond to stereo problems, the 2 following rows are multiview reconstruction instances, the last row is a shape fitting problem.

---

[6] Available at https://vision.cs.uwaterloo.ca/data/maxflow.

### 4.4   MAP Inference with Potts Potentials

Coordinate-wise minimization for the dual LP relaxation of MAP inference was intensively studied, see e.g. the review [29]. One of the formulations is

$$\min \sum_{i \in V} \max_{k \in K} \theta_i^\delta(k) + \sum_{\{i,j\} \in E} \max_{k,l \in K} \theta_{ij}^\delta(k,l) \tag{15a}$$

$$\delta_{ij}(k) \in \mathbb{R} \ \ \forall\{i,j\} \in E, k \in K, \tag{15b}$$

where $K$ is the set of labels, $V$ is the set of nodes and $E$ is the set of unoriented edges and

$$\theta_i^\delta(k) = \theta_i(k) - \sum_{j \in N_i} \delta_{ij}(k) \tag{16a}$$

$$\theta_{ij}^\delta(k,l) = \theta_{ij}(k,l) + \delta_{ij}(k) + \delta_{ji}(l) \tag{16b}$$

are equivalent transformations of the potentials. Notice that there are $2 \cdot |E| \cdot |K|$ variables, i.e. two for each direction of an edge. In [24], it is mentioned that in case of Potts interactions, which are given as $\theta_{ij}(k,l) = -[\![k \neq l]\!]$, one can add constraints

$$\delta_{ij}(k) + \delta_{ji}(k) = 0 \qquad\qquad \forall\{i,j\} \in E, k \in K \tag{17a}$$

$$-\tfrac{1}{2} \leq \delta_{ij}(k) \leq \tfrac{1}{2} \qquad\qquad \forall\{i,j\} \in E, k \in K \tag{17b}$$

to (15) without changing the optimal objective. One can therefore use constraint (17a) to reduce the overall amount of variables by defining

$$\lambda_{ij}(k) = -\delta_{ij}(k) = \delta_{ji}(k) \tag{18}$$

subject to $\tfrac{1}{2} \leq \lambda_{ij}(k) \leq \tfrac{1}{2}$. The decision of whether $\delta_{ij}(k)$ or $\delta_{ji}(k)$ should have the inverted sign depends on the chosen orientation $E'$ of the originally undirected edges $E$ and is arbitrary. Also, given values $\delta$ satisfying (17), it holds for any edge $\{i,j\} \in E$ and pair of labels $k,l \in K$ that $\max_{k,l \in K} \theta_{ij}^\delta(k,l) = 0$, which can be seen from the properties of the Potts interactions.

Therefore, one can reformulate (15) into

$$\min \sum_{i \in V} \max_{k \in K} \theta_i^\lambda(k) \tag{19a}$$

$$-\tfrac{1}{2} \leq \lambda_{ij}(k) \leq \tfrac{1}{2} \ \ \forall(i,j) \in E', k \in K, \tag{19b}$$

where the equivalent transformation in $\lambda$ variables is given by

$$\theta_i^\lambda(k) = \theta_i(k) + \sum_{(i,j) \in E'} \lambda_{ij}(k) - \sum_{(j,i) \in E'} \lambda_{ji}(k) \tag{20}$$

and we optimize over $|E'| \cdot |K|$ variables $\lambda$, the graph $(V, E')$ is the same as graph $(V, E)$ except that each edge becomes oriented (in arbitrary direction). The way

of obtaining an optimal solution to (15) from an optimal solution of (19) is given by (18) and depends on the chosen orientation of the edges in $E'$. Also observe that $\theta_i^\delta(k) = \theta_i^\lambda(k)$ for any node $i \in V$ and label $k \in K$ and therefore the optimal values will be equal. This reformulation therefore maps global optima of (19) to global optima of (15). However, it does not map interior local minima of (19) to interior local minima of (15) when $|K| \geq 3$, an example of such case is shown in Appendix D in [10].

In problems with two labels ($K = \{1, 2\}$), problem (19) is subsumed by (4) and satisfies the conditions imposed by Theorem 1 because one can rewrite the criterion by observing that

$$\max_{k \in \{1,2\}} \theta_i^\lambda(k) = \max\{\theta_i^\lambda(1) - \theta_i^\lambda(2), 0\} + \theta_i^\lambda(2) \tag{21}$$

and each $\lambda_{ij}(k)$ is present only in $\theta_i^\lambda(k)$ and $\theta_j^\lambda(k)$. Thus, $\lambda_{ij}(k)$ will have non-zero coefficient in the matrix $B$ only on columns $i$ and $j$. The coefficients of the variables in the criterion are only $\{-1, 0, 1\}$ and the other conditions are straightforward.

We reported the experiments on the Potts problem in [30] where the optimality was not proven yet. In addition, the update designed in [30] *ad hoc* becomes just a special case of our general update here.

### 4.5   Binarized Monotone Linear Programs

In [13], integer linear programs with at most two variables per constraint were discussed. It was also allowed to have 3 variables in some constraints if one of the variables occurred only in this constraint and in the objective function. Although the objective function in [13] was allowed to be more general, we will restrict ourselves to linear criterion function. It was also shown that such problems can be transformed into binarized monotone constraints over binary variables by introducing additional variables whose amount is defined by the bounds of the original variables, such optimization problem reads

$$\min w^T x + e^T z \tag{22a}$$
$$Ax - Iz \leq 0 \tag{22b}$$
$$Cx \leq 0 \tag{22c}$$
$$x \in \{0, 1\}^{n_1} \tag{22d}$$
$$z \in \{0, 1\}^{n_2}, \tag{22e}$$

where $A, C$ contain exactly one $-1$ per row and exactly one $1$ per row and all other entries are zero, $I$ is the identity matrix. We refer the reader to [13] for details, where it is also explained that the LP relaxation of (22) can be solved by min-$st$-cut on an associated graph. We can notice that the LP relaxation of (22) is subsumed by the dual (5), because one can change the minimization into maximization by changing the signs in $w, e$. Also, the relaxation satisfies the conditions given by Theorem 1.

In the paper [13], there are listed many problems which are transformable to (22) and are also directly (without any complicated transformation) subsumed by the dual (5) and satisfy Theorem 1, for example, minimizing the sum of weighted completion times of precedence-constrained jobs (ISLO formulation in [9]), generalized independent set (forest harvesting problem in [14]), generalized vertex cover [15], clique problem [15], Min-SAT (introduced in [17], LP formulation in [13]).

For each of these problems, it is easy to verify the conditions of Theorem 1, because they contain at most two variables per constraint and if a constraint contains a third variable, then it is the only occurrence of this variable and the coefficients of the variables in the constraints are from the set $\{-1, 0, 1\}$.

The transformation presented in [13] can be applied to partial Max-SAT and vertex cover to obtain a problem in the form (22) and solve its LP relaxation. But this step is unnecessary when applying the presented coordinate-wise minimization approach.

## 5    Concluding Remarks

We have presented a new class of linear programs that are exactly solved by coordinate-wise minimization. We have shown that dual LP relaxations of several well-known combinatorial optimization problems (partial Max-2SAT, vertex cover, minimum $st$-cut, MAP inference with Potts potentials and two labels, and other problems) belong, possibly after a reformulation, to this class. We have shown experimentally (in this paper and in [30]) that the resulting methods are reasonably efficient for large-scale instances of these problems. When the assumptions of Theorem 1 are relaxed (e.g., general Max-SAT instead of Max-2SAT, or the Potts problem with any number of labels), the method experimentally still provides good local (though not global in general) minima.

We must admit, though, that the practical impact of Theorem 1 is limited because the presented dual LP relaxations satisfying its assumptions can be efficiently solved also by other approaches. Thus, max-flow/min-$st$-cut can be solved (besides well-known combinatorial algorithms such as Ford-Fulkerson) by message-passing methods such as TRW-S. Similarly, the Potts problem with two labels is tractable and can be reduced to max-flow. In general, all considered LP relaxations can be reduced to max-flow, as noted in §4.5. Note, however, that this does not make our result trivial because (as noted in §2) equivalent reformulations of problems may not preserve interior local minima and thus message-passing methods are not equivalent in any obvious way to our method.

It is open whether there are practically interesting classes of linear programs that are solved exactly (or at least with constant approximation ratio) by (block-) coordinate minimization and are not solvable by known combinatorial algorithms such as max-flow. Another interesting question is which reformulations in general preserve interior local minima and which do not.

Our approach can pave the way to new efficient large-scale optimization methods in the future. Certain features of our results give us hope here. For

instance, our approach has an important novel feature over message-passing methods: it applies to a *constrained* convex problem (the box constraints (4b) and (4c)). This can open the way to a new class of applications. Furthermore, updates along large variable blocks (which we have not explored) can speed algorithms considerably, e.g., TRW-S uses updates along subtrees of a graphical model, while max-sum diffusion uses updates along single variables.

# References

1. Ansotegui, C., Bacchus, F., Järvisalo, M., Martins, R., et al.: MaxSAT Evaluation 2017 (2017), https://helda.helsinki.fi/bitstream/handle/10138/233112/mse17proc.pdf
2. Bacchus, F., Järvisalo, M., Martins, R.: MaxSAT Evaluation 2018: New developments and detailed results. Journal on Satisfiability, Boolean Modeling and Computation **11**(1), 99–131 (2019), instances available at https://maxsat-evaluations.github.io/.
3. Bacchus, F., Järvisalo, M.J., Martins, R., et al.: MaxSAT Evaluation 2018 (2018), https://helda.helsinki.fi/bitstream/handle/10138/237139/mse18_proceedings.pdf
4. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific, Belmont, MA, 2nd edn. (1999)
5. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine learning **3**(1), 1–122 (2011)
6. Boykov, Y., Lempitsky, V.S.: From photohulls to photoflux optimization. In: Proceedings of the British Machine Conf. vol. 3, p. 27. Citeseer (2006)
7. Boykov, Y., Veksler, O., Zabih, R.: Markov random fields with efficient approximations. In: Proceedings. 1998 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (Cat. No. 98CB36231). pp. 648–655. IEEE (1998)
8. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision **40**(1), 120–145 (2011)
9. Chudak, F.A., Hochbaum, D.S.: A half-integral linear programming relaxation for scheduling precedence-constrained jobs on a single machine. Operations Research Letters **25**(5), 199–204 (1999)
10. Dlask, T., Werner, T.: A class of linear programs solvable by coordinate-wise minimization. arXiv.org (2020)
11. Fulkerson, D., Ford, L.: Flows in networks. Princeton University Press (1962)
12. Globerson, A., Jaakkola, T.: Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In: Neural Information Processing Systems. pp. 553–560 (2008)
13. Hochbaum, D.S.: Solving integer programs over monotone inequalities in three variables: A framework for half integrality and good approximations. European Journal of Operational Research **140**(2), 291–321 (2002)

14. Hochbaum, D.S., Pathria, A.: Forest harvesting and minimum cuts: a new approach to handling spatial constraints. Forest Science **43**(4), 544–554 (1997)
15. Hochbaum, D.S., Pathria, A.: Approximating a generalization of MAX 2SAT and MIN 2SAT. Discrete Applied Mathematics **107**(1-3), 41–59 (2000)
16. 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. Journal of Computer Vision **115**(2), 155–184 (2015)
17. Kohli, R., Krishnamurti, R., Mirchandani, P.: The minimum satisfiability problem. SIAM Journal on Discrete Mathematics **7**(2), 275–283 (1994)
18. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE Trans. on Pattern Analysis and Machine Intelligence **28**(10), 1568–1583 (2006)
19. Kolmogorov, V.: A new look at reweighted message passing. IEEE Trans. on Pattern Analysis and Machine Intelligence **37**(5) (May 2015)
20. Kolmogorov, V., Zabih, R.: Computing visual correspondence with occlusions via graph cuts. Tech. rep., Cornell University (2001)
21. Kovalevsky, V.A., Koval, V.K.: A diffusion algorithm for decreasing the energy of the max-sum labeling problem (approx 1975), Glushkov Institute of Cybernetics, Kiev, USSR. Unpublished
22. Lempitsky, V., Boykov, Y.: Global optimization for shape fitting. In: 2007 IEEE Conf. on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (2007)
23. Lempitsky, V., Boykov, Y., Ivanov, D.: Oriented visibility for multiview reconstruction. In: European Conf. on Computer Vision. pp. 226–238. Springer (2006)
24. Průša, D., Werner, T.: LP relaxation of the Potts labeling problem is as hard as any linear program. IEEE Trans. on Pattern Analysis and Machine Intelligence **39**(7), 1469–1475 (2017)
25. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Intl. Journal of Computer Vision **47**(1-3), 7–42 (2002)
26. Schlesinger, M.I., Antoniuk, K.: Diffusion algorithms and structural recognition optimization problems. Cybernetics and Systems Analysis **47**, 175–192 (2011)
27. Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C.: A comparative study of energy minimization methods for markov random fields with smoothness-based priors. IEEE Trans. on Pattern Analysis and Machine Intelligence **30**(6), 1068–1080 (2008)
28. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. Journal of Optimization Theory and Applications **109**(3), 475–494 (2001)
29. Werner, T.: A linear programming approach to max-sum problem: A review. IEEE Trans. on Pattern Analysis and Machine Intelligence **29**(7), 1165–1179 (2007)
30. Werner, T., Průša, D., Dlask, T.: Relative interior rule in block-coordinate minimization. Accepted to IEEE/CVF Conf. on Computer Vision and Pattern Recognition (2020)
31. Werner, T., Průša, D.: Relative interior rule in block-coordinate minimization. arXiv.org (2019)
32. Wright, S.J.: Coordinate descent algorithms. Mathematical Programming **151**(1), 3–34 (2015)
33. Xu, K., Li, W.: Many hard examples in exact phase transitions with application to generating hard satisfiable instances. arXiv.org (2003), instances available at http://sites.nlsde.buaa.edu.cn/~kexu/benchmarks/max-sat-benchmarks.htm.