Mathematical Programming manuscript No.
(will be inserted by the editor)

Alexandre Belloni · Claudia Sagastizábal

Dynamic Bundle Methods

Application to Combinatorial Optimization

Dedicated to Clovis G. Gonzaga on the ocassion of his 60^{th} birthday

Abstract. Lagrangian relaxation is a popular technique to solve difficult optimization problems. However, the applicability of this technique depends on having a relatively low number of hard constraints to dualize. When there are exponentially many hard constraints, it is preferable to relax them dynamically, according to some rule depending on which multipliers are active. For instance, only the most violated constraints at a given iteration could be dualized. From the dual point of view, this approach yields multipliers with varying dimensions and a dual objective function that changes along iterations. We discuss how to apply a bundle methodology to solve this kind of dual problems. We analyze the convergence properties of the resulting dynamic bundle method, including finite convergence for polyhedral problems, and report numerical experience on Linear Ordering and Traveling Salesman Problems.

1. Introduction

Consider the following optimization problem:

$$\begin{cases} \max_p C(p) \\ p \in \mathcal{Q} \subset \mathbb{R}^{m_p} \\ g_j(p) \leq 0\,, j \in L := \{1, \ldots, n\}, \end{cases} \tag{1}$$

where $C : \mathbb{R}^{m_p} \to \mathbb{R}$, for each $j \in L$, $g_j : \mathbb{R}^{m_p} \to \mathbb{R}$ denotes a "hard" constraint, difficult to deal with. Easy constraints are included in the set $\mathcal{Q}$, which can be discrete. Suppose in addition that n is a huge integer, say bigger than 10^{10}.

Solutions to (1) can be found by using Lagrangian Relaxation techniques. Let x_j denote the nonnegative multipliers associated to hard constraints and denote the non negative orthant by $\mathbb{R}^n_{\geq 0} := \{x \in \mathbb{R}^n : x_j \geq 0 \text{ for all } 1 \leq j \leq n\}$. The dual problem of (1) is given by

$$\min_{x \in \mathbb{R}^n_{\geq 0}} f(x)\,, \qquad \text{where} \qquad f(x) := \max_{p \in \mathcal{Q}} \left\{ C(p) - \sum_{j \in L} x_j g_j(p) \right\} \tag{2}$$

is the dual function, possibly nondifferentiable.

Alexandre Belloni: Operations Research Center MIT, E40-149, 77 Massachusetts Av., Cambridge, Massachusetts 02139-4307, US. e-mail: `belloni@mit.edu`

Claudia Sagastizábal: IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro RJ 22460-320, Brazil. On leave from INRIA Rocquencourt, France. Research supported by CNPq Grant No. 303540-03/6, e-mail: `sagastiz@impa.br`

For the Lagrangian relaxation approach to make sense, two points are fundamental. First, the dual function should be much simpler to evaluate (at any given x) than solving the primal problem directly; we assume this is the case for (1). Second, the dual problem should not be large-scale: in nonsmooth optimization (NSO) this means less than 10^6 variables. So the approach is simply not applicable in our setting, because in (1) n is too big. Instead of dualizing all the n hard constraints at once, an alternative approach is to choose at each iteration *subsets* of constraints to be dualized. In this dynamical relaxation, subsets J have cardinality $|J|$ much smaller than n. As a result, the corresponding dual function, defined on $\mathbb{R}^{|J|}$, is manageable from the NSO point of view.

The idea of replacing a problem with difficult feasible set by a sequence of problems with simpler constraints can be traced back to the cutting-planes methods [CG59, Kel60]. An important matter for preventing subproblems from becoming too difficult is how and when to drop constraints. In Semi-infinite Programming, for example, the identification of *sufficiently violated* inequalities is crucial for producing implementable outer approximation algorithms, [GP79, GPT80]. In Combinatorial Optimization, a method to dynamically insert and remove inequalities in order to strengthen relaxations was already used to solve Steiner Problems in Graphs in [Luc92]. This is a special class of cutting-planes method, that was named *Relax-and-Cut* algorithm in [EGM94]. Related works are [Bea90, CB83, LB96, HFK01, MLM00, CLdS01, BL02b]. All these works use a subgradient type method, [Erm66, Sho85, HWC74] to update the dual multipliers. Although numerical experience shows the efficiency of the Relax-and-Cut technique, no convergence proof for the algorithm was given so far.

In this paper we consider Relax-and-Cut methods from a broader point of view, more focused on a dual perspective. Our procedure applies to general problems, not only combinatorial problems. Primal information produced by a *Separation Oracle* (which identifies constraints in (1) that are not satisfied, i.e., "violated inequalities") is used at each iteration to choose the subset J. Dual variables, or multipliers, are updated using a particular form of bundle methods, [HUL93, BGLS03], that we call *Dynamic Bundle Method* and is specially adapted to the setting. Thanks to the well known descent and stability properties of the bundle methodology, we are in a position to prove convergence of our method.

We mention [Hel01], a closely related work, where a similar algorithm is applied to solve semidefinite relaxations of combinatorial problems. The dual step therein uses the so-called spectral bundle method, while primal information is given by a *maximum violation oracle* which seeks for the most violated inequality. Our Separation Oracle, depending on an index set J, is slightly more general; see Remark 1 below.

This paper is organized as follows. In Section 2, the general Relax-and-Cut formulation is discussed. Sections 3 and 4 describe the bundle methodology in this new context. In Sections 5 and 6 we state the formal proofs of convergence, including finite convergence for polyhedral problems. Numerical results for Linear Ordering and Traveling Salesman problems are shown in Section 7. Finally, in Section 8 we give some concluding remarks.

Our notation and terminology is standard in bundle methods; see [BGLS03], [HUL93]. We use the euclidean inner product in $\mathbb{R}^n$: $\langle x, y\rangle = \sum_{j=1}^n x_j y_j$, with induced norm denoted by $|\cdot|$. For an index set J, $|J|$ stands for its cardinality. The indicator function of the non negative orthant $\mathbb{R}^n_{\geq 0}$, denoted by $\mathrm{I}_{\geq 0}$, is defined to be 0 for all $x \in \mathbb{R}^n_{\geq 0}$ and $+\infty$ otherwise. Finally, given $\tilde{x}$ nonnegative, $N_{\mathbb{R}^n_{\geq 0}}(\tilde{x})$ is the usual normal cone of Convex Analysis:

$$N_{\mathbb{R}^n_{\geq 0}}(\tilde{x}) = \left\{\nu \in \mathbb{R}^n : \langle \nu, x - \tilde{x}\rangle \leq 0 \text{ for all } x \in \mathbb{R}^n_{\geq 0}\right\}.$$

2. Combining primal and dual information

Relax-and-cut methods use primal information to choose which constraints in (1) will be dualized at each iteration.

The dual function f from (2) is the maximum of a collection of affine functions of x, so it is convex. Moreover, assume that (1) has one strictly feasible point (a reasonable assumption, at least when $\mathcal{Q}$ is infinite), i.e. that there exists $p' \in \mathcal{Q}$ such that $g_j(p') < 0$ for all $j \in L$. Then the dual problem (2) has a solution (see for instance [HUL93, Prop. XII.3.2.3]). In addition, the well-known weak duality property

$$f(x) \geq C(p) \quad \text{for all } x \geq 0 \text{ and } p \text{ primal feasible} \tag{3}$$

implies in this case that f is bounded from below. Finally, note that the evaluation of f for a given value of x gives straightforwardly a subgradient. More precisely, letting $p_x \in \mathcal{Q}$ be a maximizer for which $f(x) = C(p_x) - \sum_{j\in L} x_j g_j(p_x)$, it holds that

$$-g(p_x) = -\left(g_j(p_x)_{j\in L}\right) = -(g_j(p_x),_{\{j=1,\ldots,n\}}) \in \partial f(x)\,. \tag{4}$$

Based on this somewhat minimal information, *black-box methods* generate a sequence $\{x^k\}$ converging to a solution $\bar{x}$ of the dual problem (2); see [BGLS03, Chs. 7-9]. If there is no duality gap (for example if the data in (1) is "convex enough"), the corresponding $p_{\bar{x}}$ solves (1). Otherwise, it is necessary to recover a primal solution with some heuristic technique. Such techniques usually make use of $p_{\bar{x}}$ or even of the primal iterates p_{x^k}; see for example [BA00], [BMS02].

An important consequence of considering a subset J instead of the full L is that complete knowledge of dual function is no longer available. Namely, for any given x only

$$\max_{p\in\mathcal{Q}}\left\{C(p) - \sum_{j\in\mathbf{J}} x_j g_j(p)\right\} \quad \textbf{and not} \quad \max_{p\in\mathcal{Q}}\left\{C(p) - \sum_{j\in\mathbf{L}} x_j g_j(p)\right\}$$

is known. In order to obtain convergent algorithms in the absence of complete data, primal and dual information should be combined adequately. We first explain how to select hard constraints using primal points.

2.1. Selecting inequalities

To choose which constraints are to be dualized at each iteration, we assume that a *Separation Oracle* is available.

A *Separation Oracle* is a procedure $SepOr$ that, given $p \in \mathbb{R}^{m_p}$ and $J \subseteq L$, identifies inequalities in $L \backslash J$ violated by p (in other words, hard constraints in (1) that are not satisfied by p). The output of the procedure is an index set I, i.e., $I = SepOr(p, J)$, which can be empty.

We assume that, as long as there remain inequalities in $L \backslash J$ violated by p, the Separation Procedure is able to identify one of such constraints. With this assumption,

$$SepOr(p, J) = \emptyset \Leftrightarrow \{j \in L : g_j(p) > 0\} \subseteq J \Leftrightarrow \{j \in L : g_j(p) \leq 0\} \supseteq L \backslash J. \quad (5)$$

In particular, $SepOr(p, L)$ is always the empty set. In Combinatorial Optimization this assumption corresponds to having an *exact separation* algorithm for cutting the hard constraints. An efficient algorithm for such separation may not be known for many families of valid inequalities for some NP-hard problems (for example, the path inequalities for the Traveling Salesman Problem; see Section 7.2 below).

Remark 1. The Separation Oracle in [Hel01] does not make use of an index set J, and corresponds to $J = \emptyset$ in our setting:

$$SepOr_{[\mathrm{Hel01}]}(p) = SepOr(p, \emptyset).$$

In [Hel01, Def. 4.1], to guarantee that the separation procedure does not stall, but explores all the inequalities, the following "maximum violation oracle" condition

$$SepOr_{[\mathrm{Hel01}]}(p) \subseteq \left\{ j \in L : g_j(p) = \max_{l \in L} \{g_l(p)\} > 0 \right\}$$

is assumed. In particular, $SepOr_{[\mathrm{Hel01}]}(p)$ only answers the empty set if p is feasible in (1).

Even though stated for sets J, our assumption can be related to the one in [Hel01], because each set J is a subset of L. More precisely, a Separation Oracle that finds all violated inequalities by exploring the whole set L can as well explore any given subset J to decide if there remain violated inequalities. Reciprocally, a Separation Oracle for which (5) holds for all J, satisfies in particular the maximum violation condition for $J = L$. Note, however, that it is possible to build an artificial example [1] such that the only one active constraint at the solution will never be the most violated inequality, for any iteration. Thus, the Separation Oracle proposed in [Hel01] may fail to generate the right inequalities.

[1] Given a positive parameter K, consider the problem

$$\max_{p \in [0,2]^2} \left\{ p_1 + p_2 : g_1(p) = p_1 - 1 \leq 0, g_2(p) = p_2 - 1 \leq 0, g_3(p) = \frac{2}{K}(p_1 + p_2) - \frac{1}{K} \leq 0 \right\}$$

and suppose the three constraints are relaxed. At the optimal set $\{(p_1, \frac{1}{2} - p_1) : p_1 \in [0, \frac{1}{2}]\}$, g_3 is the only active constraint. Since the relaxed problem is linear, for any multiplier x the evaluation of the dual function will return as p_x a extreme point in $[0,2]^2$. So $g_i(p) \in \{-1, 1\}$

We now consider in detail which is the dual data available at each iteration.

2.2. Dual information

Given an index set $J \subset L = \{1, \ldots, n\}$, write $\mathbb{R}^n = \mathbb{R}^{|J|} \times \mathbb{R}^{n-|J|}$. We denote by $\mathcal{P}_J$ the linear operator in $\mathbb{R}^n$ defined as the orthogonal projection on the corresponding subspace:

$$\begin{aligned} \mathcal{P}_J : \mathbb{R}^n &\longrightarrow \mathbb{R}^{|J|} \times \mathbb{R}^{n-|J|} \\ x \quad &\longmapsto \left(x_J, 0_{L\setminus J}\right) := \left(x_{j\in J}, 0_{j\in L\setminus J}\right). \end{aligned}$$

With this definition, it holds that

$$x = \mathcal{P}_J(x) \quad \Longrightarrow \quad x = \mathcal{P}_{J'}(x) \text{ for all } J' \supseteq J\,. \tag{6}$$

As we mentioned, at each iteration only a *partial knowledge* of the dual function f is available. Namely, the composed function $f \circ \mathcal{P}_J$, where the index set J varies along iterations. From [HUL93, Thm. VI.4.2.1], the corresponding subdifferential is given by the formula

$$\partial(f \circ \mathcal{P}_J)(x) = \mathcal{P}_J^*\big(\partial f(\mathcal{P}_J(x))\big) = \mathcal{P}_J\big(\partial f(\mathcal{P}_J(x))\big), \tag{7}$$

because $\mathcal{P}_J^* = \mathcal{P}_J$.

NSO methods use dual data to define affine minorants of the dual function. When all the constraints are dualized at once, given $\Big(x^i, f(x^i), s^i \in \partial f(x^i)\Big)$, by the subgradient inequality, the relation $f(x^i) + \langle s^i, x - x^i\rangle \le f(x)$ always holds. In our dynamic setting, there is also an index set J_i. If we used (7) to compute subgradients, we would only get minorizations of the form

$$f\big(\mathcal{P}_{J_i}(x^i)\big) + \big\langle s\big(\mathcal{P}_{J_i}(x^i)\big), x - \mathcal{P}_{J_i}(x^i)\big\rangle \le f(x)\,,$$

which only hold on the subspace $\{x \in \mathbb{R}^n : x = \mathcal{P}_{J_i}(x)\}$. For a similar relation to hold at any x, one should rather define subgradients omitting the projection operation, as shown by the next result.

Lemma 2. *With the notation and definitions above, given $x^i \in \mathbb{R}^n_{\ge 0}$ and $J_i \subset L$, let p^i be a maximizer defining $f_i := f(\mathcal{P}_{J_i}(x^i))$:*

$$p^i \in \operatorname{Arg}\max_{p\in\mathcal{Q}} \left\{ C(p) - \sum_{j\in J_i} x^i_j g_j(p) \right\} \tag{8}$$

and define $s^i := -g(p^i) = -\left(g_j(p^i)_{j\in L}\right)$. Then

$$f_i + \big\langle s^i, x - \mathcal{P}_{J_i}(x^i)\big\rangle = C(p^i) - \sum_{j\in L} x_j g_j(p^i) \le f(x) \quad \textit{for all } x \in \mathbb{R}^n\,. \tag{9}$$

---

for $i = 1, 2$ and $|g_3(p)| \le \frac{7}{K}$. Finally, if $g_1(p) \le 0, g_2(p) \le 0$, this implies that $g_3(p) \le 0$. Thus, in order for the third constraint to be violated, at least one of the previous two constraints must also be violated. As a result, for any $K > 7$, the third constraint will never be the most violated constraint.

Proof. Use the definition of the dual function f in (2) and the definition of the projection $\mathcal{P}_{J_i}$ to write:

$$\begin{aligned} f_i &= C(p^i) - \sum_{j\in J_i} x^i_j g_j(p^i) \\ &= C(p^i) - \sum_{j\in L} \mathcal{P}_{J_i}(x^i)_j g_j(p^i)\,. \end{aligned}$$

From the definition of s^i, the first equality in (9) follows. As for the inequality in (9), it follows too from the definition of f, since it implies that $f(x) \geq C(p^i) - \sum_{j\in L} x_j g_j(p^i)$. □

By Lemma 2, the affine function $f_i + \left\langle s^i, \cdot - \mathcal{P}_{J_i}(x^i)\right\rangle$ is a minorant of the dual function f from (2). Bundle methods are based on the iterative construction of a *model* of f, defined as the maximum of planes tangent to *graph* f. Each plane is given by an affine function of the form $f_i + \left\langle s^i, \cdot - x^i\right\rangle$. Since our dynamic algorithm defines iterates x^i satisfying $x^i = \mathcal{P}_{J_i}(x^i)$, by (9), the model never cuts off a section of *graph* f. Hence, replacing f by its model does not result in loss of information, and the method eventually finds a minimum in (2) if the sets J_i are properly defined.

For convenience, in Algorithm 4 below $(f_i, s^i, p^i) := DualEval(x^i, J_i)$ denotes the dual computations described by Lemma 2.

3. Bundle methods

To help getting a better understanding of how the dynamic bundle method works, we emphasize the main modifications that are introduced by the dynamic dualization scheme when compared to a standard bundle algorithm.

Unless required for clarity, we drop iteration indices in our presentation: $\hat{x}$ and $\hat{x}^+$ below stand for certain current and next iteration stability centers, respectively.

Since (2) is a constrained NSO problem, there is an additional complication in the algebra commonly used in bundle methods. We assume that the primal problem (1) is such that the dual function f is finite everywhere. In this case, bundle methods are essentially the same than for unconstrained NSO, because the dual problem in (2) is equivalent to the unconstrained minimization of $(f + \mathrm{I}_{\geq 0})$.

3.1. An overview of bundle methods

Let ℓ denote the current iteration of a bundle algorithm. Classical bundle methods keep memory of the past in a *bundle* of information consisting of

$$\left\{f_i := f(x^i)\,, s^i \in \partial f(x^i)\right\}_{i\in\mathcal{B}} \quad \text{and } (\hat{x}, f(\hat{x})) \text{ the last “serious” iterate}\,.$$

Serious iterates, also called stability centers, form a subsequence $\{\hat{x}^k\} \subset \{x^i\}$ such that $\{f(\hat{x}^k)\}$ is strictly decreasing. Although not explicitly denoted, the superscript k in $\hat{x}^k$ depends on the current iteration, i.e., $k = k(\ell)$.

When not all the hard constraints are dualized, there is no longer a fixed dual function f, but dual objectives $f \circ \mathcal{P}_J$, with J varying along iterations. For example, $\hat{J}$ below denotes the index set used when generating the stability center $\hat{x}$ and, similarly, J_i corresponds to x^i. With the notation from Lemma 2, the bundle data is

$$\left\{ f_i := f(p^i)\,, s^i := -\big(g_j(p^i)_{j\in L}\big) \right\}_{i\in\mathcal{B}} \text{ and } \big(\hat{x}, f(\hat{x}), \hat{J}\big) \text{ with } \hat{x} = \mathcal{P}_{\hat{J}}(\hat{x}).$$

We will see that, by construction, it always holds that $x^i = \mathcal{P}_{J_i}(x^i)$. Using the (nonnegative) linearization errors

$$e_i := f(\hat{x}) - f_i - \left\langle s^i, \mathcal{P}_{\hat{J}}(\hat{x}) - \mathcal{P}_{J_i}(x^i)\right\rangle = f(\hat{x}) - f_i - \left\langle s^i, \hat{x} - x^i\right\rangle ,$$

yields the bundle of information:

$$\left\{ e_i\,, s^i = -\big(g_j(p^i)_{j\in L}\big) \right\}_{i\in\mathcal{B}} \text{ and } (\hat{x}, f(\hat{x}), \hat{J}) \text{ with } \hat{x} = \mathcal{P}_{\hat{J}}(\hat{x})\,. \tag{10}$$

Past information is used to define at each iteration a *model* of the dual function f, namely the cutting-planes function

$$\check{f}_{\mathcal{B}}(\hat{x} + d) = f(\hat{x}) + \max_{i\in\mathcal{B}} \left\{ -e_i + \left\langle s^i, d\right\rangle \right\} ,$$

where the bundle $\mathcal{B}$ varies along iterations.

A well known property of bundle methods is that it is possible to keep in the bundle $\mathcal{B}$ any number of elements without impairing convergence; see for instance [BGLS03, Ch. 9]. More precisely, it is enough for the new bundle $\mathcal{B}^+$ to contain the last generated information and the so-called *aggregate* couple $(\hat{e}, \hat{s})$; see Lemma 3 below.

An iterate x^ℓ is considered good enough to become the new stability center when $f(x^\ell)$ provides a significant decrease, measured in terms of the nominal decrease δ_ℓ

$$\delta_\ell := f(\hat{x}) - \check{f}_{\mathcal{B}}(x^\ell).$$

When an iterate is declared a serious iterate, it becomes the next stability center and $\hat{x}^+ = x^\ell$ (otherwise, $\hat{x}^+ = \hat{x}$). Linearization errors are then updated, using a recursive formula.

3.2. Defining iterates in a dynamic setting

We now consider more in detail the effect of working with a dynamic dualization scheme.

Let J be an index set such that $\hat{J} \subseteq J$ (so that we have $\hat{x} = \mathcal{P}_J(\hat{x})$, by (6)). To define an iterate satisfying $x^\ell = \mathcal{P}_J(x^\ell)$, we solve a quadratic programming problem (QP) depending on a varying set J and on a positive parameter μ. More specifically, $d^\ell = (d^\ell_J, 0_{L\setminus J})$ solves:

$$\begin{cases} \min \check{f}_{\mathcal{B}}(\hat{x}+d) + \frac{1}{2}\mu|d|^2 \\ d = \mathcal{P}_J(d) \in \mathbb{R}^n \\ \hat{x}_j + d_j \geq 0 \text{ for } j \in J\,. \end{cases}$$

The next iterate is given by $x^\ell := \hat{x} + d^\ell$.

Here arises a major difference with a purely static bundle method. Instead of just solving $\min_d \left(\check{f}_{\mathcal{B}} + \mathrm{I}_{\geq 0}\right)(\hat{x}+d) + \frac{1}{2}\mu|d|^2$, we introduce the additional constraint $d = \mathcal{P}_J(d)$ which depends on the index set J. We use the notation $\mathbb{R}^n_{J,\geq 0} := \{x \in \mathbb{R}^n_{\geq 0} : x = \mathcal{P}_J(x)\}$ for such constraint set.

Before giving the algorithm we derive some technical relations that are important for the convergence analysis. For convenience, in the QP (12) below, instead of $\check{f}_{\mathcal{B}}$ we use a more general function φ.

Lemma 3. *Let $\hat{x} \in \mathbb{R}^n_{\geq 0}$ be such that $\hat{x} = \mathcal{P}_{\hat{J}}(\hat{x})$. For $i \in I$, let $e_i \geq 0$ and $s^i \in \mathbb{R}^n$ define the function*

$$\varphi(x) := f(\hat{x}) + \max_{i\in I}\left\{\left\langle s^i, x - \hat{x}\right\rangle - e_i\right\}, \tag{11}$$

and consider the problem

$$\min_{d_J \in \mathbb{R}^{|J|}} \left\{ \left(\varphi + \mathrm{I}_{\geq 0}\right)\left(\hat{x} + (d_J, 0_{L\setminus J})\right) + \frac{1}{2}\mu|d_J|^2 \right\} \tag{12}$$

where $\hat{J} \subset J \subset L$ and $\mu > 0$. The following hold:

(i) A dual problem for (12) is

$$\begin{cases} \min\limits_{\alpha} \dfrac{1}{2\mu} \sum\limits_{j\in J} \left(\min(0, \sum\limits_{i\in I} \alpha_i s^i_j) \right)^2 + \sum\limits_{i\in I} \alpha_i e_i \\ \alpha \in \Delta_I := \{z \in \mathbb{R}^{|I|} : z_i \geq 0\,, \sum_{i\in I} z_i = 1\}. \end{cases} \tag{13}$$

(ii) Let d^*_J and α^* denote, respectively, the solutions to (12) and (13), and define $\nu^*_J := -\max(0, \sum_{i\in I}\alpha^*_i s^i_J)$. Then $x^* := \hat{x} + (d^*_J, 0_{L\setminus J})$ satisfies $x^* = \mathcal{P}_J(x^*)$ and

$$x^* = \hat{x} - \frac{1}{\mu}(\hat{s} + \hat{\nu}) \quad \textit{with} \quad \begin{cases} \hat{s} := \sum_{i\in I}\alpha^*_i s^i \in \partial\varphi(x^*) \\ \hat{\nu} := (\nu^*_J, -\sum_{i\in I}\alpha^*_i s^i_{L\setminus J}) \in N_{\mathbb{R}^n_{J,\geq 0}}(x^*). \end{cases} \tag{14}$$

Furthermore, $\varphi(x^*) = f(\hat{x}) + \left\langle s^i, x^* - \hat{x}\right\rangle - e_i$ for all $i \in I$ such that $\alpha^*_i > 0$.

(iii) The normal element satisfies $\hat{\nu} \in N_{\mathbb{R}^n_{\geq 0}}(x^*) \iff \sum_{i\in I}\alpha^*_i s^i_{L\setminus J} \geq 0$, $\mathcal{P}_J(\hat{\nu}) \in N_{\mathbb{R}^n_{\geq 0}}(x^*)$, and $\langle \hat{\nu}, \hat{x} - x^*\rangle = 0$. The nonnegative quantity $\hat{e} := \sum_{i\in I}\alpha^*_i e_i$ satisfies $\hat{e} = f(\hat{x}) - \varphi(x^*) - \mu|x^* - \hat{x}|^2$.

Suppose, in addition, that $f(x) \geq \varphi(x)$ *for all* $x \in \mathbb{R}^n_{\geq 0}$*. Then*

(*iv*) $\hat{s} + \mathcal{P}_J(\hat{\nu}) \in \partial_{\hat{e}}\left(f + \mathrm{I}_{\mathbb{R}^n_{\geq 0}}\right)(\hat{x})$*. The same inclusion holds for* $\hat{s} + \hat{\nu}$ *whenever* $\hat{\nu} \in N_{\mathbb{R}^n_{\geq 0}}(x^*)$*.*

(*v*) *The nominal decrease* $\delta = f(\hat{x}) - \varphi(x^*)$ *satisfies the relations* $\delta = \hat{e} + \frac{1}{\mu}|\hat{s} + \hat{\nu}|^2$ *and* $\delta = \Delta + \frac{1}{2}\mu|x^* - \hat{x}|^2$*, where* Δ *is the optimal value in* (13)*.*

Proof. Rewrite (12) and associate multipliers to the constraints as follows

$$\begin{cases} \min\limits_{r\in\mathbb{R}, d_J \in \mathbb{R}^{|J|}} r + \dfrac{1}{2}\mu|d_J|^2 \\ r \geq f(\hat{x}) + \left\langle s^i, (d_J, 0_{L\setminus J})\right\rangle - e_i \, , i \in I \quad (\leftrightarrow \alpha_i) \\ d_J \geq 0 \quad (\leftrightarrow -\nu_J)\, . \end{cases}$$

The corresponding KKT system is

$$\begin{cases} \mu d_J + \sum_{i\in I} \alpha_i^* s_J^i + \nu_J = 0 & \text{(a)} \\ \alpha \in \Delta_I & \text{(b)} \\ \forall i \in I \;\; \alpha_i(f(\hat{x}) + \left\langle s^i, (d_J, 0_{L\setminus J})\right\rangle - e_i - r) = 0 & \text{(c)} \\ \forall j \in J \quad \nu_j \leq 0 \quad \text{and} \quad \nu_j d_j = 0. & \text{(d)} \end{cases}$$

Let $\gamma := \sum_{i\in I} \alpha_i s^i$. Using the expression for d_j from (a) in the complementarity condition (d), we obtain for ν_j the system

$$\nu_j \leq 0 \qquad \text{and} \qquad \nu_j(\gamma_j + \nu_j) = 0,$$

solved by $\nu_J = -\max(0, \gamma_j)$. As a result,

$$\sum_{i\in I} \alpha_i^* s_J^i + \nu_J^* = \min(0, \sum_{i\in I} \alpha_i^* s_J^i),$$

so (a) rewrites as $\mu d_j + \min(0, \sum_{i\in I} \alpha_i^* s_J^i) = 0$. Use this relation to replace d_J in the Lagrangian

$$L(r, d_J, \alpha, \nu_J) = \frac{1}{2}\mu|d_J|^2 + f(\hat{x}) + \left(\sum_{i\in I} \alpha_i^* s_J^i + \nu_J\right)^\top d_J - \sum_i \alpha_i e_i$$

and drop the constant term $f(\hat{x})$ to obtain the objective function in (13).
To show item (*ii*), note first that $x^* = \mathcal{P}_J(x^*)$ because $d_J^* \geq 0$ and, since $\hat{J} \subseteq J$, $\mathcal{P}_J(\hat{x}) = \hat{x}$ by (6). Subgradients in $\partial\varphi(x^*)$ are convex combinations of s^i, for those indices $i \in I$ where the maximum in (11), written with $x = x^*$, is attained. Thus, from (b) and (c) written for α^*, d_J^*, and $r = \varphi(x^*)$, we see that $\hat{s} \in \partial\varphi(x^*)$. The last assertion in item (*ii*) follows from (c). Finally, the expression for $\hat{\nu}$ is obtained from (d), using the following characterization of normal elements ν in the cone $N_{\mathbb{R}^n_{J,\geq 0}}(x^*)$:

$$\begin{cases} \nu_j = 0 \;\; \text{for } j \in J \text{ such that } x_j^* > 0 \\ \nu_j \leq 0 \;\; \text{for } j \in J \text{ such that } x_j^* = 0 \\ \nu_j \in \mathbb{R} \text{ for } j \in L\setminus J; \end{cases}$$

see for instance [HUL93, Ex. III.5.2.6(b)]. In particular, when $J = L$, the two first assertions in item (*iii*) follow. Together with the complementarity condition (d), we obtain $\langle \hat\nu, \hat x - x^*\rangle = \nu_J^{*\top} d_J^* - (\sum_{i\in I} \alpha_i^* s^i_{L\setminus J})^\top 0_{L\setminus J} = 0$. Positivity of $\hat e$ follows from the fact that α_i^* and e_i are nonnegative for all $i \in I$. The expression for $\hat e$ in (*iii*) results from equating the primal optimal value

$$\varphi(x^*) + \frac{1}{2}\mu|x^* - \hat x|^2$$

with $f(\hat x) - \Delta$, the dual optimal value, i.e., with

$$f(\hat x) - \frac{1}{2\mu}|\sum_{i\in I} \alpha_i^* s^i_J + \nu_J^*|^2 - \hat e,$$

and using the identity $\mu^2|x^* - \hat x|^2 = |\hat s + \hat\nu|^2 = |\sum_{i\in I} \alpha_i^* s^i_J + \nu_J^*|^2$ from (14). To obtain the subgradient inclusion, we first show that

$$\langle \hat s + \vartheta, \hat s + \hat\nu\rangle = |\hat s + \hat\nu|^2 \quad \text{where } \vartheta \text{ denotes either } \mathcal{P}_J(\hat\nu) \text{ or } \hat\nu. \tag{15}$$

The relation is straighforward in the latter case. When $\vartheta = \mathcal{P}_J(\hat\nu)$, separate the J and $L\setminus J$ components of $\hat\nu$ and use its definition from (14) to write

$$\hat\nu = \mathcal{P}_J(\hat\nu) + \mathcal{P}_{L\setminus J}(\hat\nu) = \vartheta + \Big(0_J, -\sum_{i\in I} \alpha_i^* s^i_{L\setminus J}\Big).$$

Hence, $\hat s + \hat\nu = \hat s + \vartheta + \Big(0_J, -\sum_{i\in I} \alpha_i^* s^i_{L\setminus J}\Big)$ and, therefore,

$$|\hat s + \hat\nu|^2 = \langle \hat s + \vartheta, \hat s + \hat\nu\rangle + \left\langle \Big(0_J, -\sum_{i\in I} \alpha_i^* s^i_{L\setminus J}\Big), \hat s + \hat\nu \right\rangle = \langle \hat s + \vartheta, \hat s + \hat\nu\rangle\,,$$

because $(\hat s + \hat\nu)_{L\setminus J} = 0$ by (14).

Since we suppose that in both cases $\vartheta \in N_{\mathbb{R}^n_{\geq 0}}(x^*) = \partial I_{\geq 0}(x^*)$, for item (*iv*) to hold, the $\hat e$-subgradient inequality

$$(f + \mathrm{I}_{\geq 0})(x) \geq f(\hat x) + \langle \hat s + \vartheta, x - \hat x\rangle - \hat e$$

must be satisfied by each $x \in \mathbb{R}^n$. By definition of indicator function, the relation holds trivially for all $x \notin \mathbb{R}^n_{\geq 0}$. Let $x \in \mathbb{R}^n_{\geq 0}$; the facts that $\hat s \in \partial\varphi(x^*)$ and ϑ is a normal element give the inequalities

$$\begin{array}{rl} \varphi(x) \geq \varphi(x^*) + & \langle \hat s, x - x^*\rangle \\ 0 \geq & \langle \vartheta, x - x^*\rangle\,. \end{array}$$

By adding these inequalities and using that $f(x) \geq \varphi(x)$, we see that

$$\begin{aligned} f(x) &\geq \varphi(x^*) + \langle \hat s + \vartheta, x - x^*\rangle \\ &= f(\hat x) \; + \langle \hat s + \vartheta, x - \hat x\rangle - (f(\hat x) - \varphi(x^*) - \langle \hat s + \vartheta, \hat x - x^*\rangle). \end{aligned}$$

The desired result follows from the expression for $\hat{e}$ and (15).
Finally, item (v) follows the identity $\mu^2|x^* - \hat{x}|^2 = |\hat{s} + \hat{\nu}|^2$ and the expression for $\hat{e}$ in item (iii). □

The aggregate couple $(\hat{e}, \hat{s})$ from Lemma 3 represents in a condensed form the information related to active bundle elements, i.e., to $I^{act} := \{i \in I : \alpha_i^* > 0\}$. Along the iterative bundle process, this couple may be inserted in the index set I defining φ for the next iteration. Namely, when $|I^{act}|$ becomes too big and some active elements must be discarded. When the aggregate couple enters the index set I, the corresponding aggregate primal point may be useful to separate inequalities, cf. $\hat{\pi}^\ell$ in Corollary 6(i) below.

In the algorithm stated in section 4 below, φ is the cutting-plane model $\check{f}_{\mathcal{B}}$, $I = \mathcal{B}$ and $s^i = -g(p^i)$. Stability centers are indexed by k. Both the bundle $\mathcal{B}$ and the index set J may vary with the iteration index ℓ, and the same holds for the QP information, where we use the notation

$$\left(x^\ell, \hat{s}^\ell, \hat{\nu}^\ell, \hat{e}_\ell, \delta_\ell, {\alpha^*}^\ell, \{p^i : i \in \mathcal{B}^{act}\}\right) = QPIt(\hat{x}^k, \check{f}_{\mathcal{B}}, J_\ell, \mu_\ell)\,.$$

4. The algorithm

All the elements are now in place for the precise algorithm to be defined.

Algorithm 4 (Dynamic Bundle Method).

Initialization. *Choose parameters* $m \in (0, 1]$, $\mu_0 > 0$, $|\mathcal{B}|_{max} \geq 2$, *and a tolerance* $tol \geq 0$. *Choose* $0 \neq x^0 \in \mathbb{R}^n_{\geq 0}$ *and define* $J_0 := \{j \leq n : x_j^0 > 0\}$.
Make the dual computations to obtain $(f_0, s^0, p^0) := DualEval(x^0, J_0)$.
Set $k = 0$, $\hat{x}^0 := x^0$, $\hat{f}_0 := f_0$, *and* $\hat{J}_0 := J_0$. *Set* $\ell = 1$ *and let* $J_1 := J_0$; *define the aggregate bundle* $\hat{\mathcal{B}} := \emptyset$ *and the oracle bundle* $\mathcal{B} := \{(e_0 := 0, s^0)\}$.

Step 1. (Iterate Generation) *Solve the QP problem:*

$$\left(x^\ell, \hat{s}^\ell, \hat{\nu}^\ell, \hat{e}_\ell, \delta_\ell, {\alpha^*}^\ell, \{p^i : i \in \mathcal{B}^{act}\}\right) = QPIt(\hat{x}^k, \check{f}_{\mathcal{B}}, J_\ell, \mu_\ell)\,.$$

Step 2. (Dual Evaluation) *Make the dual computations*

$$(f_\ell, s^\ell, p^\ell) := DualEval(x^\ell, J_\ell)\,.$$

Step 3. (Descent test and Separation)
If $f_\ell \leq \hat{f}_k - m\delta_\ell$ *then declare a* serious step*:*
Move the stabilization center: $\left(\hat{x}^{k+1}, \hat{f}_{k+1}\right) := \left(x^\ell, f_\ell\right)$.
Update the linearization and aggregate errors:

$$e_i = e_i + \hat{f}_{k+1} - \hat{f}_k - \left\langle s^i, \hat{x}^{k+1} - \hat{x}^k \right\rangle \text{ for } i \in \mathcal{B}$$
$$\hat{e}_\ell = \hat{e}_\ell + \hat{f}_{k+1} - \hat{f}_k - \left\langle \hat{s}^\ell, \hat{x}^{k+1} - \hat{x}^k \right\rangle$$

Compute $O_{k+1} := \{j \in J_\ell : \hat{x}_j^{k+1} = 0\}$ *and define* $\hat{J}_{k+1} := J_\ell \backslash O_{k+1}$.
Call the separation procedure to compute

$$I_\ell := \bigcup \left\{ SepOr(p^i, \hat{J}_{k+1}) : i \in \mathcal{B}^{act} \right\}.$$

Let $J_{\ell+1} := \hat{J}_{k+1} \cup I_\ell$. *Set* $k = k+1$.

Otherwise, declare a **null step***: compute* $e_\ell := \hat{f}_k - f_\ell - \langle s^\ell, \hat{x}^k - x^\ell \rangle$.

Call the separation procedure to compute

$$I_\ell := \bigcup \{ SepOr(p^i, J_\ell) : i \in \mathcal{B}^{act} \}.$$

Set $J_{\ell+1} := J_\ell \cup I_\ell$.

Step 4. (Stopping test) *If* $\delta_\ell \le tol$ *and* $I_\ell = \emptyset$ *stop.*

Otherwise go to Step 5.

Step 5. (Bundle Management and Update)

If the bundle has not reached the maximum bundle size, set $\mathcal{B}_{red} := \mathcal{B}$.

Otherwise, if $|\mathcal{B}| = |\mathcal{B}|_{max}$, *then delete at least one element from* $\mathcal{B}$ *to obtain* $\mathcal{B}_{red}$. *If* $\mathcal{B}^{act} \subset \mathcal{B}_{red}$, *then* $\hat{\mathcal{B}} = \emptyset$. *If* $\mathcal{B}^{act} \not\subset \mathcal{B}_{red}$, *then delete at least one more element from* $\mathcal{B}_{red}$ *and replace* $\hat{\mathcal{B}}$ *by* $\hat{\mathcal{B}} := \{(\hat{e}_\ell, \hat{s}^\ell)\}$.

In all cases define $\mathcal{B}^+ := \mathcal{B}_{red} \cup \hat{\mathcal{B}} \cup \{(e_\ell, s^\ell)\}$.

Loop *Choose* $\mu_{\ell+1} > 0$. *Set* $\mathcal{B} = \mathcal{B}^+$, $\ell = \ell + 1$ *and go to Step 1.* □

Before passing to the convergence analysis, we comment on some features of the algorithm.

- Note that J_0 is defined in order to satisfy $x^0 = \mathcal{P}_{J_0}(x^0)$. Since by construction, it always holds that $J_\ell \supset \hat{J}_{k(l)}$, Lemma 3($i$) ensures that $x^\ell = \mathcal{P}_{J_\ell}(x^\ell)$ for all subsequent $\ell > 0$.
- Index sets defined at serious steps could also be defined by $\hat{J}_k = \{j \in L : \hat{x}^k_j > 0\} \cup I_\ell$, where $I_\ell \subset \{j \in L : \hat{x}^k_j = 0 \text{ and } g_j(p^k) > 0\}$.
- Step 5 selects bundle elements trying to keep all active bundle elements. If this is not possible, the aggregate couple is inserted in the bundle.

5. Convergence Analysis. Dual Results

We will show in Section 6 that a variant of Algorithm 4 (which keeps all active elements so that $\hat{\mathcal{B}}_\ell$ is empty for all ℓ) has finite termination whenever $\mathcal{Q}$ is finite and the Assumption 11 given below holds. We proceed in two steps, similarly to [Kiw85, Ch. 2.6]. Namely, we first show that if the algorithm loops forever, it converges to a solution of a certain dual problem. Then we use this asymptotic result to obtain a contradiction and prove finite termination in Section 6.

In this section we focus on dual convergence results. For our presentation we mostly follow [HUL93, Ch. XV.3]. Notwithstanding the formal similarities with the refered material, the dynamic scheme introduces modifications in the QP that need to be addressed carefully.

5.1. The effect of compressing the bundle

Associated to the aggregate couple that may be inserted in the bundle at Step 5 to replace deleted active elements, there is an affine function, sometimes called *aggregate linearization.*

Lemma 5. *With the notation and definitions of Algorithm 4, consider the following affine function, defined at the ℓ^{th}-iteration for all $x \in \mathbb{R}^n$:*

$$\tilde{f}_\ell(x) := \check{f}_{\mathcal{B}}(x^\ell) + \left\langle \hat{s}^\ell, x - x^\ell \right\rangle \ . \tag{16}$$

The following holds:

(i) $\tilde{f}_\ell \leq \check{f}_{\mathcal{B}}$ *and* $\tilde{f}_\ell(x) = \hat{f}_k + \left\langle \hat{s}^\ell, x - \hat{x}^k \right\rangle - \hat{e}_\ell$.

(ii) At each iteration of Algorithm 4, it holds that

$$\check{f}_{\mathcal{B}}(x) \leq f(x) \quad \text{for all } x \in \mathbb{R}^n. \tag{17}$$

(iii) Let $\mathcal{B}^+$ denote the updated bundle defined at the end of Step 5 in Algorithm 4. Then $\check{f}_{\mathcal{B}^+}(x^\ell) = f(x^\ell) = f_\ell$ and $\check{f}_{\mathcal{B}^+}(x^{\ell+1}) \geq \tilde{f}_\ell(x^{\ell+1})$.

In particular, all the statements in Lemma 3 hold, written for $\hat{x} = \hat{x}^k$, $x^* = x^\ell$, $\hat{J} = \hat{J}_k$, $J = J_\ell$, $I = \mathcal{B}_\ell$, $\varphi = \check{f}_{\mathcal{B}}$, $\mu = \mu_\ell$, $\hat{s} = \hat{s}^\ell$, $\hat{\nu} = \hat{\nu}^\ell$, $\hat{e} = \hat{e}_\ell$, and $\delta = \delta_\ell$.

Proof. Items $(i) - (iii)$ in Lemma 3 hold without any assumption, the remaining statements, i.e., items (iv) and (v), require that $\varphi \leq f$ on $\mathbb{R}^n_{\geq 0}$, an inequality that will follow from (17).

The first relation in item (i) is straightforward from (16), because $\hat{s}^\ell \in \partial \check{f}_{\mathcal{B}}(x^\ell)$ by (14). To show the second relation, add and substract $\left\langle \hat{s}^\ell, \hat{x}^k \right\rangle$ to the right hand side of (16). Then use the expression for $\hat{e}_\ell$ in Lemma 3 (iii) and the relation $\mu_\ell \left\langle \hat{s}^\ell, \hat{x}^k - x^\ell \right\rangle = |\hat{s}^\ell + \hat{\nu}^\ell|^2$, which follows from the expression for x^ℓ in (14) and the last assertion in Lemma 3(ii).

To see item (ii), first note that, by definition of $\tilde{f}_\ell$, $\nabla \tilde{f}_\ell(x) = \hat{s}^\ell$ for all x, in particular for $x = x^\ell$. Thus, the optimality condition (14) is also the optimality condition for a QP as in (12), with φ replaced by $\tilde{f}_\ell$. Now let Φ be any function satisfying

$$\tilde{f}_\ell(x) \leq \Phi(x) \text{ for each } x \in \mathbb{R}^n \text{ and } \Phi(x^\ell) = \tilde{f}_\ell(x^\ell). \tag{18}$$

Since $\Phi \geq \tilde{f}_\ell$ with equality at x^ℓ, we see the iterate x^ℓ is also the solution to the QP (12), with φ replaced by Φ.

Consider now $\mathcal{B}^+$, the updated bundle defined at the end of Step 5 in Algorithm 4. By construction, it has the form $\mathcal{B}^+ = \mathcal{B}_{red} \cup \hat{\mathcal{B}} \cup \{(e_\ell, s^\ell)\}$, so $\check{f}_{\mathcal{B}^+}(x) = \max\{\Phi(x), f_\ell + \left\langle s^\ell, x - x^\ell \right\rangle\}$ where the expression for $\Phi := \check{f}_{\mathcal{B}_{red} \cup \hat{\mathcal{B}}}$ has three possibilites. We claim that in all cases Φ satisfies the conditions in (18). When $\hat{\mathcal{B}} = \emptyset$, either $\mathcal{B}_{red} = \mathcal{B}$ or $\mathcal{B}_{red} \supset \mathcal{B}^{act}$. In the first case, $\Phi := \check{f}_{\mathcal{B}}$, and the claim holds by item (i). In the second case, $\Phi := \check{f}_{\mathcal{B}_{red}}$. Clearly, $\check{f}_{\mathcal{B}^{act}} \geq \tilde{f}_\ell$ and, since $\mathcal{B}_{red} \supset \mathcal{B}^{act}$, $\check{f}_{\mathcal{B}_{red}} \geq \check{f}_{\mathcal{B}^{act}}$. In addition, because $\mathcal{B} \supset \mathcal{B}_{red}$, $\check{f}_{\mathcal{B}}(x^\ell) \geq \check{f}_{\mathcal{B}_{red}}(x^\ell) \leq \check{f}_{\mathcal{B}^{act}}(x^\ell)$. But from the last assertion in Lemma 3(ii), $\check{f}_{\mathcal{B}}(x^\ell) = \hat{f}^k + \left\langle s^i, \hat{x}^k - x^\ell \right\rangle - e_i$ for all $i \in \mathcal{B}^{act}$. So $\Phi(x^\ell) = \check{f}_{\mathcal{B}_{red}}(x^\ell) = \tilde{f}_\ell(x^\ell)$. Finally, when $\hat{\mathcal{B}} \neq \emptyset$, $\Phi := \max\{\check{f}_{\mathcal{B}_{red}}, \tilde{f}_\ell\}$, so $\Phi \geq \tilde{f}_\ell$. In particular, $\Phi(x^\ell) \geq \tilde{f}_\ell(x^\ell) = \check{f}_{\mathcal{B}}(x^\ell)$, by (16). In addition, by item (i), $\tilde{f}_\ell \leq \check{f}_{\mathcal{B}}$ and, since $\mathcal{B}_{red} \subset \mathcal{B}$, $\check{f}_{\mathcal{B}_{red}} \leq \check{f}_{\mathcal{B}}$. It follows tat $\Phi \leq \check{f}_{\mathcal{B}}$. In particular, $\Phi(x^\ell) \leq \check{f}_{\mathcal{B}}(x^\ell)$, which shows that Φ satisfies the relations in (18) and our claim is proved for all three cases.

To show item (ii), we proceed by induction on ℓ, the iteration counter of Algorithm 4. In our notation, $\mathcal{B} = \mathcal{B}_\ell$ and $\mathcal{B}^+ = \mathcal{B}_{\ell+1}$. By (9) written for x^0, (17) holds for the starting bundle, because $\check{f}_{\mathcal{B}_1}(x) = f(x^0) + \langle s^0, x - x^0 \rangle$ and f is convex. Suppose (17) is satisfied for $\mathcal{B}$. We have that $\check{f}_{\mathcal{B}^+}(x) = \max\{\Phi(x), f_\ell + \langle s^\ell, x - x^\ell \rangle\}$, $\Phi \le \check{f}_{\mathcal{B}}$. Since by the inductive assumption $\Phi(x) \le f(x)$ for all $x \in \mathbb{R}^n$ and by (9) the affine piece defined from (e_ℓ, s^ℓ) is a minorant for f, (17) holds for $\mathcal{B}^+$.
Finally, to see item (iii), first note that $\check{f}_{\mathcal{B}^+}(x^\ell) = \max\{\Phi(x^\ell), f_\ell\}$, with $\Phi(x^\ell) \le \check{f}_{\mathcal{B}}(x^\ell)$. Since $\check{f}_{\mathcal{B}}(x^\ell) \le \tilde{f}_\ell$, the first assertion in item (iii) follows. The second assertion follows from a similar reasoning and using (18), which yields that $\check{f}_{\mathcal{B}^+}(x^{\ell+1}) \ge \Phi(x^{\ell+1}) \ge \tilde{f}_\ell(x^{\ell+1})$. □

Lemma 5 above ensures that the compression mechanism in Step 5 of Algorithm 4 does not introduce any change into dual iterations, at least from the global convergence point of view. In practice, however, it is observed that small bundle sizes, say $|\mathcal{B}|_{\max} = 2$, do affect speed of convergence. Even though each QP resolution at Step 2 is faster, the total number of iterations increases and quite often the total CPU time is higher than when using "reasonable" values of $|\mathcal{B}|_{\max}$ (preferably, keeping all active elements).

However, depending on the original problem (1), compression of primal information may not be possible. Algorithm 4 makes use of active primal points in order to compute a new index set I_ℓ. Thus, a *primal bundle* of past primal points p^i (computed in Step 2 at the i^{th}-iteration) should also be kept along iterations. Such primal bundle may be compressed without any loss of information, following rules similar to those in Step 5. We now express some of the relations in Lemma 3 in a primal form. In particular, the equivalence result stated in item (iii) therein reveals which are the primal points that should be separated by the Separation Oracle.

Corollary 6. *Recall that $\mathcal{B}^{act} = \{i \in \mathcal{B} : \alpha_i^{*^\ell} > 0\}$ denotes the active simplicial indices in Step 1 of Algorithm 4. Suppose that in (1) g is affine and $\mathcal{Q}$ is convex. The following relations hold:*

(i) *The subgradient in (14) has the form $\hat{s}^\ell = -g(\hat{\pi}^\ell)$, where*

$$\hat{\pi}^\ell := \sum_{i \in \mathcal{B}^{act}} \alpha_i^{*^\ell} p^i \qquad \textit{for certain } p^i \in \mathcal{Q}.$$

(ii) *If $SepOr(\hat{\pi}^\ell, J_\ell) = \emptyset \Longrightarrow \hat{\nu} \in N_{\mathbb{R}^n_{\ge 0}}(x^\ell)$.*

(iii) *If C is also affine, and e_i in (11) satisfies $e_i = f(\hat{x}^k) - f(x^i) - \langle s^i, \hat{x}^k - x^i \rangle$ for all $i \in \mathcal{B}^{act}$, then*

$$\check{f}_{\mathcal{B}}(x^\ell) = C(\hat{\pi}^\ell) - \langle g(\hat{\pi}^\ell), x^\ell \rangle \tag{19}$$

and x^ℓ, $\check{f}_{\mathcal{B}}(x^\ell)$ and α^{*^ℓ} solve the KKT system

$$\begin{cases} \forall j \in J_\ell \quad \mu(x_j^\ell - \hat{x}_j^k) + \min(0, \sum_{i \in I} \alpha_i^* s_j^i) = 0 & \text{(a)} \\ \alpha^{*^\ell} \in \Delta_{\mathcal{B}} & \text{(b)} \\ \forall i \in \mathcal{B}^{act} \ \alpha_i^{*^\ell}(C(p^i) - \langle g(p^i), x^* \rangle - \check{f}_{\mathcal{B}}(x^\ell)) = 0 & \text{(c)} \end{cases} \tag{20}$$

corresponding to (12) *with* $\varphi = \check{f}_{\mathcal{B}^{act}}$.

Proof. By (14), $\hat{s} = \sum_{i\in\mathcal{B}^{act}} \alpha_i^{*^\ell} s^i$. By construction, each s^i is either an aggregate $\hat{s}$, or s^i satisfies $s^i = -g(p^i) \in \partial f(x^i)$ with x^i and p^i given by Lemma 2. When $s^i = \hat{s}$, again by (14), it is the convex sum of past subgradients, so eventually $s^i = -\sum_{j\le i} \beta_j g(p^j)$ for some $p^j \in \mathcal{Q}$. Item (i) follows, because g is affine. To show (ii), suppose that the set $SepOr(\hat{\pi}^\ell, J_\ell)$ is empty. Then by (5) $g_j(\hat{\pi}^\ell) \le 0$ for all $j \in L\backslash J_\ell$, and the first assertion in Lemma 3(iii) gives the desired inclusion, because $\sum_{i\in\mathcal{B}^{act}} \alpha_i^{*^\ell} s^i_{L\backslash J_\ell} = g_{L\backslash J_\ell}(\hat{\pi}^\ell)$.
To show (19), use, sucessively, the expression for $\hat{e}$ in Lemma 3(iii), (14), the fact that $\langle \hat{\nu}, x^* - \hat{x}\rangle = 0$ by Lemma 3(iii), and the definitions of $\hat{s}$ and $\hat{e}$ to obtain

$$\begin{aligned}\varphi(x^*) &= f(\hat{x}) + \langle \mu(\hat{x} - x^*), x^* - \hat{x}\rangle - \hat{e}\\ &= f(\hat{x}) + \langle \hat{s} + \hat{\nu}, x^* - \hat{x}\rangle - \hat{e}\\ &= f(\hat{x}) + \langle \hat{s}, x^* - \hat{x}\rangle - \hat{e}\\ &= f(\hat{x}) + \textstyle\sum_{i\in I^{act}} \alpha_i^* \left(\langle s^i, x^* - \hat{x}\rangle - e_i\right).\end{aligned}$$

By definition of e_i, $f(\hat{x}) - \langle s^i, \hat{x}\rangle - e_i = f(x^i) - \langle s^i, x^i\rangle$, so (19) follows from (9), and similarly for (20), recalling the expression for the KKT system associated to (12). □

We see from these results that when g is affine and $\mathcal{Q}$ convex, to define I_ℓ in Step 3 of Algorithm 4 it is enough to separate the convex primal point $\hat{\pi}^\ell$. Otherwise, we may be compelled to stock all active primal points. Finally, note that if I_ℓ is empty for all ℓ, then the condition on linearization errors in Corollary 6(iii) holds and, thus, both (19) and (20) apply.

5.2. Asymptotic results

When the algorithm loops forever, there are two mutually exclusive situations. Either there is a last serious step followed by an infinite number of null steps, or there are infinitely many different serious steps.

We give first with a technical result, for a particular instance of iterates, namely, two consecutive null steps with same index set J_ℓ.

Corollary 7. *With the notation and definitions of Algorithm 4, let* ℓ *be an iteration index giving a null step. Suppose the next iteration* $\ell+1$ *also gives a null step. If* $\mu_\ell \le \mu_{\ell+1}$ *and* $J_{\ell+1} = J_\ell$*, then* Δ_ℓ*, the optimal value of* (13)*, satisfies the relation*

$$\Delta_{\ell+1} + \frac{1}{2}\mu_\ell |x^{\ell+1} - x^\ell|^2 \le \Delta_\ell \,.$$

Proof. First expand squares to obtain the identity

$$\frac{1}{2}\mu_\ell |x^\ell - \hat{x}^k|^2 + \frac{1}{2}\mu_\ell |x^{\ell+1} - x^\ell|^2 = \mu_\ell \left\langle \hat{x}^k - x^\ell, x^{\ell+1} - x^\ell\right\rangle + \frac{1}{2}\mu_\ell |x^{\ell+1} - \hat{x}^k|^2.$$

The scalar product in the right hand side above is equal to $\langle \hat{s}^\ell + \hat{\nu}^\ell, x^{\ell+1} - x^\ell \rangle$, by (14). Since $J_{\ell+1} = J_\ell$ by assumption and $\mathcal{P}_{J_\ell}(\hat{\nu}^\ell) \in N_{\mathbb{R}^n_{\geq 0}}(x^\ell)$ by Lemma 3(iii), we have that $\langle \hat{\nu}^\ell, x^{\ell+1} - x^\ell \rangle = \langle \hat{\nu}^\ell, \mathcal{P}_{J_\ell}(x^{\ell+1} - x^\ell) \rangle = \langle \mathcal{P}_{J_\ell}(\hat{\nu}^\ell), x^{\ell+1} - x^\ell \rangle \leq 0$ and, thus,

$$\frac{1}{2}\mu_\ell |x^\ell - \hat{x}^k|^2 + \frac{1}{2}\mu_\ell |x^{\ell+1} - x^\ell|^2 \leq \langle \hat{s}^\ell, x^{\ell+1} - x^\ell \rangle + \frac{1}{2}\mu_\ell |x^{\ell+1} - \hat{x}^k|^2.$$

Combine the last inequality with the definition of Δ_ℓ, (16), and the second assertion in Lemma 5 (iii) to see that

$$\begin{aligned} \hat{f}_k - \Delta_\ell + \frac{1}{2}\mu_\ell |x^{\ell+1} - x^\ell|^2 &\leq \check{f}_{\mathcal{B}}(x^\ell) + \langle \hat{s}^\ell, x^{\ell+1} - x^\ell \rangle + \frac{1}{2}\mu_\ell |x^{\ell+1} - \hat{x}^k|^2 \\ &= \tilde{f}_\ell(x^{\ell+1}) + \frac{1}{2}\mu_\ell |x^{\ell+1} - \hat{x}^k|^2 \\ &\leq \check{f}_{\mathcal{B}^+}(x^{\ell+1}) + \frac{1}{2}\mu_\ell |x^{\ell+1} - \hat{x}^k|^2 . \end{aligned}$$

The result follows, because the assumption that $\mu_\ell \leq \mu_{\ell+1}$ implies that $\hat{f}_k - \Delta_\ell + \frac{1}{2}\mu_\ell |x^{\ell+1} - x^\ell|^2 \leq \check{f}_{\mathcal{B}^+}(x^{\ell+1}) + \frac{1}{2}\mu_{\ell+1} |x^{\ell+1} - \hat{x}^k|^2 = \hat{f}_k - \Delta_{\ell+1}$. □

We now consider the case of infinite iterations of Algorithm 4, starting with the case of a finite number of serious steps.

Lemma 8. *Consider Algorithm 4 applied to the minimization problem* (2)*. Suppose there is an iteration $\hat{\ell}$ where the stability center $\hat{x}$ is generated, followed by an infinite number of null steps. Suppose, in addition, that $\mu_\ell \leq \mu_{\ell+1}$. Then there is an iteration $\ell ast \geq \hat{\ell}$ such that the following holds for all $\ell \geq \ell ast$:*

(i) $J_\ell = \bar{J} \subset L$.
(ii) *If, in addition, $m \in (0,1)$ then $\delta_\ell \to 0$, $x^\ell \to \hat{x}$, and $\check{f}_{\mathcal{B}}(x^\ell) \to f(\hat{x})$.*

Proof. Since we only remove inequalities at serious steps via O_{k+1}, once $\hat{x}$ has been generated, index sets J_ℓ can only increase. Since L is finite, there is eventually an iteration, say ℓast, such that $J_{\ell ast} = \bar{J} \subset L$. Subsequently, there are only null steps with $J_\ell = \bar{J}$ for all $\ell \geq \ell ast$, as stated in item (i).
To see (ii), first note that by (i) for $\ell \geq \ell ast$ Corollary 7 applies. In particular, the sequence $\{\Delta_\ell\}$ is decreasing. Using succesively the definition of Δ_ℓ, (14), (16) and the fact that $\langle \hat{\nu}^\ell, \hat{x} - x^\ell \rangle = 0$ by Lemma 3(iii), we obtain the following identities:

$$\begin{aligned} f(\hat{x}) - \Delta_\ell + \frac{1}{2}\mu_\ell |\hat{x} - x^\ell|^2 &= \check{f}_{\mathcal{B}}(x^\ell) + \mu_\ell |\hat{x} - x^\ell|^2 \\ &= \check{f}_{\mathcal{B}}(x^\ell) + \langle \hat{s}^\ell + \hat{\nu}^\ell, \hat{x} - x^\ell \rangle \\ &= \tilde{f}_\ell(\hat{x}) + \langle \hat{\nu}^\ell, \hat{x} - x^\ell \rangle \\ &= \tilde{f}_\ell(\hat{x}) . \end{aligned}$$

Together with $\tilde{f}_\ell(\hat{x}) \le \check{f}_{\mathcal{B}}(\hat{x}) \le f(\hat{x})$, by Lemma 5($i$) and (17), we obtain that $-\Delta_\ell + \frac{1}{2}\mu_\ell|\hat{x} - x^\ell|^2 \le 0$. Therefore, since $\mu_\ell \le \mu_{\ell+1}$ and $\{\Delta_\ell\}$ is decreasing, we conclude that

$$\frac{1}{2}\mu_{\ell ast}|\hat{x} - x^\ell|^2 \le \frac{1}{2}\mu_\ell|\hat{x} - x^\ell|^2 \le \Delta_\ell \le \Delta_{\ell ast}\,. \tag{21}$$

Thus, the sequence $\{x^\ell\}$ remains in a bounded set. We denote by C a Lipschitz constant for both f and $\check{f}_{\mathcal{B}^+}$ on such set. Using that $-f_\ell + \check{f}_{\mathcal{B}^+}(x^\ell) = 0$ by Lemma 5(iii),

$$f_{\ell+1} - \check{f}_{\mathcal{B}^+}(x^{\ell+1}) = f_{\ell+1} - f_\ell + \check{f}_{\mathcal{B}^+}(x^\ell) - \check{f}_{\mathcal{B}^+}(x^{\ell+1}) \le 2C|x^{\ell+1} - x^\ell|\,. \tag{22}$$

Now, for $m < 1$, sum the following inequalities (resulting, respectively, from the facts that $\Delta_{\ell+1} = \delta_{\ell+1} + \frac{1}{2}\mu_{\ell+1}|x^{\ell+1} - \hat{x}^k|^2$ and $x^{\ell+1}$ is a null step:

$$\begin{aligned} \Delta_{\ell+1} &\le f(\hat{x}) - \check{f}_{\mathcal{B}^+}(x^{\ell+1}) \\ -m\Delta_{\ell+1} \le -m\delta_{\ell+1} &\le f_{\ell+1} - f(\hat{x}) \end{aligned}$$

to obtain, together with (22), that

$$(1-m)\Delta_{\ell+1} \le f_{\ell+1} - \check{f}_{\mathcal{B}^+}(x^{\ell+1}) \le 2C|x^{\ell+1} - x^\ell|\,.$$

As a result, because μ_ℓ is non decreasing,

$$\frac{1}{2}\mu_\ell|x^{\ell+1} - x^\ell|^2 \ge \frac{(1-m)^2}{8C^2}\mu_\ell\Delta_{\ell+1}^2 \ge \frac{(1-m)^2}{8C^2}\mu_{\ell ast}\Delta_{\ell+1}^2\,,$$

which, together with Corollary 7 yields, after summation, that $\Delta_\ell \to 0$, because

$$\frac{(1-m)^2}{8C^2}\mu_{\ell ast}\sum_{\ell\ge\ell ast}\Delta_{\ell+1}^2 \le \sum_{\ell\ge\ell ast}(\Delta_\ell - \Delta_{\ell+1}) \le \Delta_{\ell ast}\,.$$

In particular $\delta_\ell \to 0$ and the third assertion in item (ii) follows. The second assertion follows from (21) and the facts that $\mu_{\ell ast} \le \mu_\ell$ and $x^\ell \to \hat{x}$. □

The next result is now straightforward.

Corollary 9. *Suppose that in* (1) g *is affine and* $\mathcal{Q}$ *is convex. With the assumptions and notation in Lemma 8, the last generated stability center* $\hat{x}$ *is a minimizer of*

$$\begin{cases} \min f(x) \\ x_j \ge 0 & \text{for } j \in \bar{J} \\ x_j = 0 & \text{for } j \in L\backslash\bar{J}. \end{cases}$$

If, in addition, $I_{\hat{\ell}} = \emptyset$, *then* $\hat{x}$ *solves* (2).

Proof. For contradiction purposes, suppose there exists $\bar{x} \geq 0$ with $\bar{x}_{L\setminus\bar{J}} = 0$ such that $f(\bar{x}) < f(\hat{x})$. Let φ be a convex function satisfying $\varphi \leq f$ and $\varphi(\hat{x}) = f(\hat{x})$. For any $s \in \partial\varphi(\hat{x})$,

$$\begin{aligned} 0 > f(\bar{x}) - f(\hat{x}) \geq \varphi(\bar{x}) - f(\hat{x}) &\geq \langle s, \bar{x} - \hat{x}\rangle \\ &= \langle s, \mathcal{P}_{\bar{J}}(\bar{x}) - \hat{x}\rangle + \langle s, \mathcal{P}_{L\setminus\bar{J}}(\bar{x})\rangle. \end{aligned} \tag{23}$$

By Corollary 6(i), $\hat{s}^\ell = -g(\hat{\pi}^\ell)$. Since by Lemma 3$(ii)$, $|x^\ell - \hat{x}|^2 = \frac{1}{\mu_\ell}|\hat{s}^\ell + \hat{\nu}^\ell|^2$, from Lemma 8 we have that $|\hat{s}^\ell + \hat{\nu}^\ell| \to 0$. For $j \in L\setminus\bar{J}$, since $I_{\hat{\ell}} = \emptyset$, $\hat{s}^\ell_j \geq 0$. If $j \in \bar{J}$ and $\hat{x}_j = 0$, then $\hat{\nu}^\ell_j \leq 0$, thus $dist(\hat{s}^\ell_j, \mathbb{R}_+) \to 0$. Finally, if $j \in \bar{J}$ and $\hat{x}_j > 0$, $\hat{\nu}^\ell_j = 0$ which implies that $\hat{s}^\ell_j \to 0$. Therefore, the right hand side in (23) written with $s = \hat{s}^\ell$ and $\varphi = \check{f}_{\mathcal{B}}$ converges to a nonnegative number, a contradiction. □

The proof of Lemma 8 puts in evidence an interesting issue. Namely, as long as $J_{\ell+1} \supset J_\ell$ at null steps, from the dual point of view it does not really matter how the index set is chosen. The situation is quite the opposite at serious steps. In this case we show that, to ensure convergence, we are bound to requiring eventual exact separation of the active primal elements.

Lemma 10. *Consider Algorithm 4 applied to the minimization problem* (2). *Suppose there there are infinitely many serious steps and let* ℓ_k *denote an iteration index giving a serious step:* $\hat{x}^{k+1} = x^{\ell_k}$. *Then*

(i) $\delta_{\ell_k} \to 0$ *when* $k \to \infty$.

Suppose that in (1) g *is affine and* $\mathcal{Q}$ *is convex. Suppose, in addition, that the Separation Procedure satisfies* (5) *and there is a descent iteration index* K *such that for all* $k \geq K$

$$I_{\ell_k} = \bigcup \left\{ SepOr(p^i, \hat{J}_{k+1}) : i \in \mathcal{B}^{act}_{\ell_k} \right\} = \emptyset\,.$$

Then, if $0 < \mu_{min} \leq \mu_{\ell_k} \leq \mu_{max}$ *for all* $k \geq K$*, the following holds:*

(ii) There exists $\bar{x}$ *solving* (2) *such that* $\{\hat{x}^k\} \to \bar{x}$ *and* $\check{f}_{\mathcal{B}_{\ell_k}}(\hat{x}^{k+1}) \to f(\bar{x})$.

(iii) $J_{\ell_k} \supset \bar{J} := \{j \in L : \bar{x}_j > 0\}$ *for all* $k \geq K$.

Proof. Let $\bar{f}$ denote the optimal value in (2), which is finite because the assumption that there exists a strictly feasible primal point implies that f is bounded below. When the descent test holds, $\hat{f}_{k+1} \leq \hat{f}_k - m\delta_{\ell_k}$ for all k. By summation over the infinite set of serious step indices, we obtain that $0 \leq m\sum_k \delta_{\ell_k} \leq \hat{f}_0 - \bar{f}$. Thus, the series of nominal decreases is convergent: $\lim_k \delta_{\ell_k} = 0$, as stated in item (i). From Lemma 3(v), both the series $\hat{e}_{\ell_k}$ and $\frac{1}{\mu_{\ell_k}}|\hat{s}^{\ell_k} + \hat{\nu}^{\ell_k}|^2$ converge:

$$\lim_k \hat{e}_{\ell_k} = 0 \quad \text{and, because } \mu_{\ell_k} \leq \mu_{max}, \quad \lim_k |\hat{s}^{\ell_k} + \hat{\nu}^{\ell_k}|^2 = 0\,. \tag{24}$$

Our assumptions and Corollary 6(ii) imply that for all $k \geq K$ the element $\hat{\nu}^{\ell_k}$ is in the normal cone. By Lemma 3(iv), $f(x) \geq \hat{f}_k + \left\langle \hat{s}^{\ell_k} + \hat{\nu}^{\ell_k}, x - \hat{x}^k \right\rangle - \hat{e}_k$, i.e.,

$$\forall x \in \mathbb{R}^n_{\geq 0} \text{ it holds that } \left\langle \hat{s}^{\ell_k} + \hat{\nu}^{\ell_k}, x - \hat{x}^k \right\rangle \leq f(x) - \hat{f}_k + \hat{e}_k \,. \tag{25}$$

In the square below, add $\pm\hat{x}^k$ and use the relation $0 = \hat{s}^{\ell_k} + \hat{\nu}^{\ell_k} + \mu_{\ell_k}(\hat{x}^{k+1} - \hat{x}^k)$ given by (14) to obtain

$$\begin{aligned} |x - \hat{x}^{k+1}|^2 &= |x - \hat{x}^k|^2 + \left\langle x - \hat{x}^k, \hat{x}^k - \hat{x}^{k+1} \right\rangle + |\hat{x}^k - \hat{x}^{k+1}|^2 \\ &= |x - \hat{x}^k|^2 + \frac{2}{\mu_{\ell_k}} \left(\left\langle x - \hat{x}^k, \hat{s}^{\ell_k} + \hat{\nu}^{\ell_k} \right\rangle + \frac{1}{2}\mu_{\ell_k} |\hat{s}^{\ell_k} + \hat{\nu}^{\ell_k}|^2 \right) . \end{aligned}$$

Together with (25), using the expression for δ_{ℓ_k} in Lemma 3(v) and the fact that $\mu_{\ell_k} \geq \mu_{min} > 0$, we have

$$|x - \hat{x}^{k+1}|^2 \leq |x - \hat{x}^k|^2 + \frac{2}{\mu_{min}} \left(f(x) - \hat{f}_k + \delta_{\ell_k} \right) \quad \text{for all } x \in \mathbb{R}^n_{\geq 0} \,. \tag{26}$$

In particular, (26) implies that the sequence $\{\hat{x}^k\}$ is minimizing for (2). Otherwise, suppose for contradiction purposes that there is $x^* \in \mathbb{R}^n_{\geq 0}$ and $\omega > 0$ such that $f(x^*) + \omega \leq \hat{f}_k$. From (26) written for $x = x^*$,

$$|x^* - \hat{x}^{k+1}|^2 \leq |x^* - \hat{x}^k|^2 + \frac{2}{\mu_{min}} \left(-\omega + \delta_{\ell_k} \right) .$$

Since $\delta_{\ell_k} \to 0$, there is eventually a $K_* \geq K$ such that for all $k \geq K^*$ it holds that $\delta_{\ell_k} \leq \omega/2$ and, thus,

$$\frac{1}{\mu_{min}} \omega \leq |x^* - \hat{x}^k|^2 - |x^* - \hat{x}^{k+1}|^2 \,.$$

By summation over such k, we obtain the desired contradiction.
Now let $\bar{x}$ be a minimizer of f on $\mathbb{R}^n_{\geq 0}$ (which exists, because (1) has a strictly feasible point, by assumption). By (26) written for $x = \bar{x}$,

$$|\bar{x} - \hat{x}^{k+1}|^2 \leq |\bar{x} - \hat{x}^k|^2 + \frac{2}{\mu_{min}} \delta_{\ell_k} \,, \tag{27}$$

which shows that the sequence $\{\hat{x}^k\}$ is bounded, because $\delta_{\ell_k} \to 0$. As a result, any cluster point of the (minimizing) sequence $\{\hat{x}^k\}$ is a minimizer to (2). Let $\bar{x}$ denote such accumulation point. For any given $\omega > 0$, let $K_\omega \geq K$ such that

$$|\bar{x} - \hat{x}^{K_\omega}|^2 \leq \frac{\omega}{2} \quad \text{and} \quad \frac{2}{\mu_{min}} \sum_{k=K_\omega}^{+\infty} \delta_{\ell_k} \leq \frac{\omega}{2} \,.$$

Let $k' \geq K_\omega$ be an arbitrary serious step index. By summation of (27) over $K_\omega \leq k \leq k'$, we obtain that

$$|\bar{x} - \hat{x}^{k'+1}|^2 \leq |\bar{x} - \hat{x}^{k_\omega}|^2 + \frac{2}{\mu_{min}} \sum_{k=K_\omega}^{k'} \delta_{\ell_k} \,.$$

Therefore, $|\bar{x}-\hat{x}^{k'+1}|^2 \leq \omega$ for all k' big enough and $\bar{x}$ is the unique accumulation point of the sequence $\{\hat{x}^k\}$. The fact that $\check{f}_{\mathcal{B}_{\ell_k}}(\hat{x}^k) \to f(\bar{x})$ follows from (26) written for $x = \bar{x}$ and the definition of δ_{l_k}.
Finally, to show item (iii) we proceed again by contradiction. If there are infinite serious step indices k' such that $\bar{J} \not\subset J_{\ell_{k'}}$, since L is finite, there must be an index $\bar{j} \in \bar{J} \backslash J_{\ell_{k'}}$. Then $\bar{x}_{\bar{j}} > 0$, but $\hat{x}^{k'}_{\bar{j}} = 0$, because $\bar{j} \notin J_{\ell_{k'}}$. Hence,

$$0 < \bar{x}_{\bar{j}} = |\hat{x}^{k'}_{\bar{j}} - \bar{x}_{\bar{j}}| \leq |\hat{x}^{k'} - \bar{x}| \,.$$

The contradiction follows, because by item (ii), $\hat{x}^k \to \bar{x}$, which means that $|\hat{x}^k - \bar{x}| \leq \omega$ for all $\omega > 0$ (for instance, $\omega := \bar{x}_{\bar{j}}/2$) and k big enough. □

In the next section we show that a variant of Algorithm 4 (which never deletes active elements from the bundle) has finite termination when the dual function is polyhedral, i.e., when $\mathcal{Q}$ is finite and the primal problem is a suitable linear program.

6. Finite termination. Primal Results

We now study when Algorithm 4 gives solutions to the primal problem (1). We consider a variant that keeps all the active elements, without ever inserting an aggregate couple in the bundle.

Primal solutions can be obtained by a dual approach only when there is no duality gap. This is the purpose of our assumption below.

Assumption 11. *We suppose that C and g, the functions defining* (1)*, are affine. Letting conv $\mathcal{Q}$ denote the convex hull of the set $\mathcal{Q}$, we suppose that the primal problem has the so-called* integrality *property, i.e., solutions to*

$$\begin{cases} \max\limits_p C(p) \\ p \in conv\, \mathcal{Q} \\ g_j(p) \leq 0 \,, j \in L, \end{cases}$$

are solutions to (1).

First we show that when $tol = 0$, if the Algorithm stops, it finds a primal optimal point.

Lemma 12. *Let Algorithm 4 be applied to the minimization problem* (2) *with $\mathcal{Q}$ replaced by conv $\mathcal{Q}$. Suppose that $\hat{\mathcal{B}}_\ell = \emptyset$ for all ℓ and there is an iteration ℓast such that $\delta_{\ell ast} = 0$. Let $k(\ell ast)$ denote the current descent index and let $\hat{\pi}^{\ell ast}$ denote the corresponding primal convex point:*

$$\hat{\pi}^{\ell ast} := \sum_{i \in \mathcal{B}^{act}_{\ell ast}} {\alpha^*}^{\ell ast} p^i.$$

If Assumption 11 holds, then $f(x^{k(\ell ast)}) = C(\hat{\pi}^{\ell ast})$.
Furthermore, if $I_{\ell ast} = SepOr(\hat{\pi}^{\ell ast}, J_{\ell ast}) = \emptyset$, then $\hat{\pi}^{\ell ast}$ is a solution to (1).

Proof. The assumption that $\hat{\mathcal{B}}_\ell = \emptyset$ implies that all items in Corollary 6 apply. By Lemma 3(v), both $\hat{e}_{\ell ast}$ and $\hat{s}^{\ell ast} + \hat{\nu}^{\ell ast}$ are zero. Therefore, by (14), $x^{\ell ast} = \hat{x}^{k(\ell ast)}$ and, by definition of δ, $f(\hat{x}^{k(\ell ast)}) = \check{f}_{\mathcal{B}_{\ell ast}}(x^{\ell ast}) = \check{f}_{\mathcal{B}_{\ell ast}}(\hat{x}^{k(\ell ast)})$. Together with (19), this gives

$$f(\hat{x}^{k(\ell ast)}) = C(\hat{\pi}^{\ell ast}) - \left\langle g(\hat{\pi}^{\ell ast}), \hat{x}^{k(\ell ast)} \right\rangle .$$

We show now that the right hand side scalar product is zero. Since $\hat{x}^{k(\ell ast)} = \mathcal{P}_{J_{\ell ast}}(\hat{x}^{k(\ell ast)})$, we only need to consider those components in $J_{\ell ast}$. By item (i) in Corollary 6, $\hat{s}_j^{\ell ast} = -g_j(\hat{\pi}^{\ell ast})$. Since $\hat{s}^{\ell ast} + \hat{\nu}^{\ell ast} = 0$, we obtain that $\hat{\nu}_j^{\ell ast} = g_j(\hat{\pi}^{\ell ast})$ for $j \in J_{\ell ast}$. Therefore, because $\mathcal{P}_{J_{\ell ast}}(\hat{\nu}^{\ell ast}) \in N_{\mathbb{R}^n_{\geq 0}}(\hat{x}^{k(\ell ast)})$, by the characterization of normal elements in [HUL93, Ex. III.5.2.6(b)], we see that

$$\begin{cases} g_j(\hat{\pi}^{\ell ast}) = 0 \text{ for } j \in \hat{J}_{k(\ell ast)} \\ g_j(\hat{\pi}^{\ell ast}) \leq 0 \text{ for } j \in J_{\ell ast} \backslash \hat{J}_{k(\ell ast)}, \end{cases}$$

so $\left\langle g(\hat{\pi}^{\ell ast}), \hat{x}^{k(\ell ast)} \right\rangle = 0$, with $g_j(\hat{\pi}^{\ell ast}) \leq 0$ for all $j \in J_{\ell ast}$. If, in addition, $SepOr(\hat{\pi}^{\ell ast}, J_{\ell ast}) = \emptyset$, then $\hat{\pi}^{\ell ast}$ is primal feasible. Together with (3), the result follows. □

We finish our analysis by studying the behaviour of Algorithm 4 for a polyhedral dual function.

6.1. *The case of a polyhedral dual function*

Suppose that $\mathcal{Q}$ in (1) is finite, with q elements $p^1, p^2, \ldots, p^q$. The corresponding dual function has the form

$$f(x) = \max_{i \leq q} \left\{ C(p^i) - \sum_{j \in L} x_j g_j(p^i) \right\} .$$

As a result, f has at most q different subgradients $s^i = -g(p^i)$. Likewise, for a fixed stability center $k(\ell)$, the linearization errors

$$e_i = f(x^{k(\ell)}) - f(x^i) - \left\langle s^i, x^{k(\ell)} - x^i \right\rangle$$

can only take a finite number of different values, at most q. We now show that in this case a specialized variant of Algorithm 4 cannot loop forever, there is an iteration ℓast such that $\delta_{\ell ast} = 0$ and, thus, Lemma 12 applies.

Theorem 13. *Consider Algorithm 4 with $m = 1$ and $\hat{\mathcal{B}}_\ell = \emptyset$ applied for solving problem* (1) *with $\mathcal{Q}$ finite. Suppose, that the Separation Procedure satisfies* (5) *and there is a descent iteration index K such that for all $k \geq K$*

$$I_{\ell_k} = \cup \left\{ SepOr(p^i, \hat{J}_{k+1}) : i \in \mathcal{B}^{act}_{\ell_k} \right\} = \emptyset .$$

Suppose in addition, that $0 < \mu_{min} \leq \mu_{\ell_k} \leq \mu_{max}$, with $\mu_\ell = \mu_{k(\ell)}$ at null step iterations. Then the algorithm stops after a finite number of iterations.

Proof. For contradiction purposes, suppose Algorithm 4 loops forever. Then the asymptotic results stated for $m \in [0,1]$ in Section 5 apply, namely, Corollary 7, item (i) in Lemma 8, and Lemma 10.
Suppose first that there is an infinite number of null steps, after a final serious step, $k(\ell ast)$. Since $\hat{\mathcal{B}}_\ell = \emptyset$, the assumption that $\mathcal{Q}$ is finite implies that for all $\ell \geq \ell ast$, linearization errors and subgradients can only have a finite number of different values, say $\{(e_1, s^1), \ldots, (e_q, s^q)\}$, even when $\ell \to \infty$. By Lemma 8(i), $J_\ell = \bar{J}$, so the QP yielding iterates x^ℓ in Step 1 eventually has the form

$$\begin{cases} \min_\alpha \dfrac{1}{2\mu_{k(\ell ast)}} \displaystyle\sum_{j \in \bar{J}} \min(0, \sum_{i \leq q} \alpha_i s^i_j)^2 + \sum_{i \leq q} \tilde{\alpha}_i e_i \\ \alpha \in \Delta := \{z \in \mathbb{R}^q : z_i \geq 0\,, \sum_{i \leq q} z_i = 1\}\,, \end{cases}$$

where the variable is no longer in $\mathbb{R}^\ell$, but in $\mathbb{R}^q$. We know from Lemma 3(v) that Δ_ℓ is the optimal value of this QP. By Corollary 7, the sequence $\{\Delta_\ell\}_{\ell > \ell ast}$ is strictly decreasing. This contradicts the fact that there is only a finite number of possible different values for $\Delta_{\ell+1}$. Therefore, there cannot be an infinite number of null steps.
Suppose now that there is an infinite number of serious steps and let ℓ_k denote an iteration giving a serious step. We consider a quadratic program without the linear term in the objective, fixing μ:

$$\begin{cases} \min_\alpha \dfrac{1}{2\mu_{\max}} \displaystyle\sum_{j \in \bar{J}} \min(0, \sum_{i \in \mathcal{B}} \alpha_i s^i_j)^2 \\ \alpha \in \Delta := \{z \in \mathbb{R}^{|\mathcal{B}|} : z_i \geq 0\,, \sum_{i \in \mathcal{B}} z_i = 1\}\,. \end{cases} \tag{28}$$

We denote by $\beta(\mathcal{B}, J)$ the optimal value of this problem. Note that there is only a finite number of different values for β, because $J \subset L$ is finite and there are at most q different subgradients s^i. Consider the particular value $\beta_{\ell_k} := \beta(\mathcal{B}^{act}_{\ell_k}, J_{\ell_k})$ and let $\bar{\alpha}$ denote the corresponding minimizer. Since $\mu_{\ell_k} \leq \mu_{max}$ and the linear term in the objective function of (13) has nonnegative values, $\beta_{\ell_k} \leq \Delta_{\ell_k}$. Thus, for ℓ_k big enough $\beta_{\ell_k} = 0$, because by Lemma 10, δ_{ℓ_k} and, hence, Δ_{ℓ_k}, go to 0. Therefore,

$$\begin{cases} \forall j \in J_{\ell_k} \ \min(0, \sum_{i \in \mathcal{B}^{act}_{\ell_k}} \bar{\alpha}_i s^i_j) = 0 & \text{(a)} \\ \qquad\qquad \bar{\alpha} \in \Delta_{\mathcal{B}} & \text{(b)} \end{cases}$$

and, for any x^* such that $\check{f}_{\mathcal{B}^{act}}(x^*) = C(p^i) - \langle g(p^i), x^* \rangle$ for all $i \in \mathcal{B}^{act}$,

$$\forall i \in \mathcal{B}^{act} \quad \bar{\alpha}_i (C(p^i) - \langle g(p^i), x^* \rangle - \check{f}_{\mathcal{B}^{act}_{\ell_k}}(x^*)) = 0 \qquad \text{(c)}.$$

In particular, $x^* = x^{\ell_k}$ satisfies the system (a)(b)(c), which by (20) gives the optimality conditions for

$$\min_{d_J \in \mathbb{R}^{|J_{\ell_k}|}} \left\{ \left(\check{f}_{\mathcal{B}^{act}_{\ell_k}} + \mathrm{I}_{\geq 0} \right) (\hat{x}^k + (d_J, 0_{L \setminus J})) \right\}.$$

Therefore, x^{ℓ_k} is one of such minimizers. We now show that the dual solution $\bar{x}$ from Lemma 10(*ii*) is also one of such minimizers. First, we show that $f(\bar{x}) = \check{f}_{\mathcal{B}_{\ell_k}}(\bar{x})$. More precisely, for all $i \in \mathcal{B}^{act}_{\ell_k}$ we have that

$$\begin{aligned} 0 \leq f(\bar{x}) - C(p^i) + \langle g(p^i), \bar{x} \rangle &= f(\bar{x}) - \check{f}_{\mathcal{B}_{\ell_k}}(x^{\ell_k}) + \check{f}_{\mathcal{B}_{\ell_k}}(x^{\ell_k}) - C(p^i) + \langle g(p^i), \bar{x} \rangle \\ &= f(\bar{x}) - \check{f}_{\mathcal{B}_{\ell_k}}(x^{\ell_k}) + \langle g(p^i), \bar{x} - x^{\ell_k} \rangle \\ &\leq f(\bar{x}) - \check{f}_{\mathcal{B}_{\ell_k}}(x^{\ell_k}) + |g(p^i)||\bar{x} - x^{\ell_k}|, \end{aligned}$$

where we used the definition of f, of $i \in \mathcal{B}^{act}_{\ell_k}$ and the Cauchy Schwarz inequality. By Lemma 10, $x^{k+1} = x^{\ell_k} \to \bar{x}$ and $\check{f}_{\mathcal{B}_{\ell_k}}(x^{k+1}) \to f(\bar{x})$ as $\ell_k \to \infty$, so the right hand side terms in the inequality above go to 0. Since the left hand side term does not depend on ℓ_k, it must be null: $f(\bar{x}) = \check{f}_{\mathcal{B}_{\ell_k}}(\bar{x}) = C(p^i) - \langle g(p^i), \bar{x} \rangle$ for all $i \in \mathcal{B}^{act}_{\ell_k}$, which means that (c) holds for $\bar{\alpha}$ and $x^* = \bar{x}$. Hence

$$f(\bar{x}) = \check{f}_{\mathcal{B}_{\ell_k}}(\bar{x}) = \check{f}_{\mathcal{B}_{\ell_k}}(x^{\ell_k})$$

for all k sufficiently large.

When $m = 1$, the descent test in Step 3 becomes $f(\hat{x}^{k+1}) \leq f(\hat{x}^k) - \delta_{\ell_k} = \check{f}_{\mathcal{B}_{\ell_k}}(\hat{x}^{k+1}) = f(\bar{x})$. Since infinite serious steps decrease the function values, $f(\bar{x}) < f(x^{k+1})$, yielding the desired contradiction. □

The various convergence results presented in Sections 5 and 6 can be summed up as follows:

– If in (1) g is affine and $\mathcal{Q}$ is convex, Algorithm 4 converges to a dual solution for all variants.
– If (1) satisfies Assumption 11 and $\mathcal{Q}$ is finite, a variant of Algorithm 4 (with $m = 1$, $\hat{\mathcal{B}}_\ell = \emptyset$, and fixed prox-parameter at null steps) stops after a finite number of iterations with a primal solution of (1).

For both results to hold, the Separation Procedure must satisfy (5) and eventually separate all inequalities for aggregate primal points $\hat{\pi}$. In order to evaluate the impact of these assumptions, we analyze the behaviour of Algorithm 4 for two Combinatorial Optimization problems.

7. Numerical results

In order to test the proposed method, two combinatorial problems were investigated. For each one, a family of inequalities was dynamically dualized, and both the Separation Oracle and the Dual Evaluation procedures were implemented. Before proceeding to the computational results, we briefly describe each problem formulation.

7.1. Linear Ordering Problems

The Linear Ordering Problem (LOP) consists in placing elements of a finite set N in sequential order. If object i is placed before object j, we incur in a cost c_{ij}. The objective is to find the order with minimum cost.

The LOP Linear Integer Programming formulation in [GJR84] uses a set of binary variables $\{p_{ij} : (i,j) \in N \times N\}$. If object i is placed before object j, $p_{ij} = 1$ and $p_{ji} = 0$.

$$\begin{cases} \min\limits_{p} \sum\limits_{(i,j):i \neq j} c_{ij} p_{ij} & \\ \quad p_{ij} + p_{ji} = 1, & \text{for every pair } (i,j) \\ \quad p_{ij} \in \{0,1\}, & \text{for every pair } (i,j) \\ \quad p_{ij} + p_{jk} + p_{ki} \leq 2, & \text{for every triple } (i,j,k) \ (*) \end{cases}$$

The 3-cycle inequalities in constraint (*) above have a huge cardinality. They are the natural candidates for our dynamic scheme, so we associate a multiplier x_{ijk} to each one of them. After relaxation, it results a concave dual function, corresponding to $-f(x)$ in (2), whose evaluation at any given x is done by solving $(N^2 - N)/2$ small problems with only two variables each. More precisely, each Dual Evaluation amounts to solving

$$\begin{cases} \sum\limits_{(i,j):i<j} \min\limits_{\{p_{ij},p_{ji}\}} \tilde{c}_{ij} p_{ij} + \tilde{c}_{ji} p_{ji} \\ \qquad p_{ij} + p_{ji} = 1 \\ \qquad p_{ij}, p_{ji} \in \{0,1\} \end{cases}$$

where $\bar{c}_{ij} = c_{ij} - \sum_{k \in N \setminus \{i,j\}} x_{ijk}$ depends on the multiplier x.

For initialization purposes, we need $J_0 \neq \emptyset$. To obtain this first set of inequalities, we just separate a solution of the relaxed problem with $J = \emptyset$. This procedure generates a non-empty set of inequalities to start with (otherwise we have found the optimal solution and we are done).

For this test problem, there is only a polynomial number of dynamically dualized constraints. The Separation Procedure is therefore easy: it just consists of checking any triple of indices, a task that can be done in constant time for each triple. So all assumptions required for dual and finite/primal convergence are met. To illustrate the application of the method we used 49 instances of the library LOLIB[2] ; see Table 1 below.

7.2. Traveling Salesman Problem

Let $G = G(V_n, E_n)$ be the undirected graph induced by a set of nodes V_n and a set of edges E_n. For each edge of the graph, a cost c_e is given, and we associate a binary variable p_e which equals one if we use the edge e in the solution and

[2] http://www.iwr.uni-heidelberg.de/groups/comopt/software/LOLIB/

equals zero otherwise. For every set $S \subset V_n$, we denote by $\delta(S) = \{(i,j) \in E_n : i \in S, j \in V_n \setminus S\}$ the edges with exactly one endpoint in S. Finally, for every set of edges $A \subset E_n$, $p(A) = \sum_{e \in A} p_e$. The Traveling Salesman Problem (TSP) formulation of Dantzig, Fulkerson and Johnson is given by

$$\begin{cases} \min\limits_{p} \sum\limits_{e \in E_n} c_e p_e \\ \quad p(\delta(\{j\})) = 2\,, \text{ for every } j \\ \quad p(\delta(S)) \geq 2\,,\ S \subset V_n, S \neq \emptyset \\ \quad p_e \in \{0,1\}\,,\ e \in E. \end{cases} \tag{29}$$

As in the well known Held and Karp (HK) bound, we make use of 1-tree structures. Let V_n denote the set of vertices, and let E_n the set of edges. Given a vertex v_1, we call a 1-tree the set of edges formed by the union of a tree on $V_n \setminus v_1$ and any two edges that have v_1 as one end node. For

$$\mathcal{X} := \{x \in \mathbb{R}^{|E_n|} : x \text{ is a 1-tree}\},$$

we define an equivalent problem of (29) by

$$\begin{cases} \min\limits_{p} \sum\limits_{e \in E_n} c_e p_e \\ \quad p(\delta(\{j\})) = 2\,, \text{ for every } j \ (\#) \\ \quad p \in \mathcal{X} \end{cases} \tag{30}$$

The HK bound is obtained by solving a dual problem resulting from dualizing constraints (30.#). In order to obtain tighter bounds, we introduce a family of facet inequalities for the associated polytope, called the r-Regular t-Paths Inequalities, following the approach in [BL02a]. More precisely, we choose vertex sets $H_1, H_2, \ldots, H_{r-1}$ and $T_1, T_2, \ldots, T_t$, called "handles" and "teeth" respectively, which satisfy the following relations:

$H_1 \subset H_2 \subset \cdots \subset H_{r-1}$
$H_1 \cup T_j \neq \phi$ for $j = 1, \ldots, t$
$T_j \setminus H_h \neq \phi$ for $j = 1, \ldots, t$
$(H_{i+1} \setminus H_i) \setminus \cup_{j=1}^{t} T_j$ for $1 \leq i \leq r-2$.

The corresponding p-Regular t-Path inequality is given by

$$\sum_{i=1}^{r-1} y(\gamma(H_i)) + \sum_{j=1}^{t} p(\gamma(T_j)) \leq \sum_{i=1}^{r-1} |H_i| + \sum_{j=1}^{t} |T_j| - \frac{t(r-1)+r-1}{2}.$$

When introducing these inequalities as additional constraints in (30), we may improve the HK bound. However, since the number of such inequalities is exponential in the size of the problem, a dynamic scheme of dualization is needed.

The initial set J_0 is given by the set of inequalities dualized by Held and Karp.

For this test problem, the Separation Oracle can only use heuristics, since no efficient exact separation procedure is known. Fortunately, the Separation Oracle is called to separate inequalities from 1-tree structures. In this case, thanks to integrality and the existence of exactly one cycle, the heuristic search for violated inequalities is much easier than for general sets. Essentially, we search first for any node with three or more edges, and then try to expand an odd number of paths from such a node.

Computational tests were made over TSPLIB[3] instances which are symmetric and have dimension smaller than 1000 cities; see Table 2 below.

7.3. Comparison with a subgradient method

To implement the dynamic version of the bundle method, we use as a starting point the code described in [Fra02]. This code is versatile enough to handle the specialized operations needed to carry out our dynamic scheme (e.g., adding/removing inequalities).

For our comparisons, we use a subgradient method, as in [BL02b] and [Luc92]. The stepsize, of the form $m(f(x^\ell)-f^*)/|s(x^\ell)|^2$ with $m \in]0,2[$, needs an estimate f^* of the optimal dual value. We use bounds given by primal feasible points in LOP problems, and (known) optimal values for TSP problems.

As usual in Combinatorial Optimization, a stopping test checking for zero duality gap is always available. If this test is not satisfied, the dual method should trigger an intrinsic stopping test, indicating that the dual function has been minimized. It is known that in general subgradient methods lack of reliable stopping tests. Hence, we stop the subgradient variants when a maximum number of iterations is reached. In Tables 1 and 2 below we denote the corresponding variant by SGnK, where n stands for a maximum of n thousands of iterations. By contrast, bundle methods do have reliable stopping tests, essentially checking when $\delta_\ell < tol$. In our runs we use $tol = 10^{-4}$ for both test problems and $|\mathcal{B}_{max}| = 25$ and 80 for LOP and TSP instances, respectively.

		Primal-Dual results		
	# DualEval	**Gap (%)**	**Time w.r.t SG3K**	
SG3K	1517	0.001	1	
Bundle	302	0.004	8.595	
		Constraints/Size (#)		
	Active	**Maximum**	**Average**	
SG3K	61.217	114.465	61.255	
Bundle	124.056	154.696	100.807	
		CPU times (%)		
	SepOracle	**DualEval**	**Manage**	**Directions**
SG3K	16.090	11.617	8.105	5.132
Bundle	5.813	4.756	14.402	74.543

Table 1. Summary of Results for LOP instances

[3] http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/

		Primal-Dual results		
	# DualEval	**Gap (%)**	**Time w.r.t. SG2K**	
SG2K	1948	0.787	1	
SG5K	4881	0.503	2.381	
Bundle	1374	0.533	2.614	
		Constraints/Size(#)		
	Active	**Maximum**	**Average**	
SG2K	0.165	0.192	0.157	
SG5K	0.158	0.190	0.154	
Bundle	1.759	2.089	0.437	
		CPU times (%)		
	SepOracle	**DualEval**	**Manage**	**Directions**
SG2K	2.434	96.906	0.321	0.192
SG5K	2.950	96.299	0.205	0.121
Bundle	3.152	36.733	7.559	52.123

Table 2. Summary of Results for TSP instances

Tables 1 and 2 summarize our average results for LOP and TSP instances. Each table contains three subtables describing primal-dual results, separation statistics, and CPU time distribution, respectively. Primal-dual results subtable reports on (average) number of dual evaluations, duality gaps, and time ratio with respect to the fastest subgradient variant. The second subtable, on the separation procedure, reports the number of active dualized constraints at the final point, as well as the maximum and average number of dualized constraints. Here, averages are taken after normalization relative to the instance size (number of objects in LOP and of cities in TSP). Finally, the third subtable reports the fraction of total CPU time spent by the Separation Oracle, by the Dual Evaluation, and by the Managing module, where we split the time spent on addition/removal of inequalities from the time needed to generate a new dual iterate.

Note that in both tables the number of Dual Evaluations required by the dynamic bundle scheme is substantially lower than in the subgradient scheme. This suggests the high quality of the directions generated by the bundle method at each iteration. Furthermore, the average number of inequalities in the Bundle Method is consistently larger, due to the “memory” imposed by our method, which only removes inequalities at serious steps. In its current version, the bundle code uses cold starts for QP subproblems. This feature, together with the fact that there are more dual variables (i.e., more dualized constraints in average), explains the huge fraction of CPU times spent on the Managing module. Warm starts (for example, starting QP matrices factorizations not from scratch, but from previous computations) should lead to a more favourable time distribution. For the Separation Oracle, we take advantage of the integrality of the relaxed primal point in order to implement more efficiently the search (or the heuristic) in the separation procedure. The small time fraction spent on the Separation Oracle reflects this feature.

8. Concluding Remarks

In this paper, we give a theoretical proof of convergence to a particular *Relax and Cut* method. The main point was to introduce a sophisticated rule for removing inequalities that ensures convergence. By combining a bundle methodology with the Relax and Cut framework, we devised Algorithm 4, and proved its convergence for the following general cases:

– If in (1) g is affine and $\mathcal{Q}$ is convex, Algorithm 4 converges to a dual solution for all variants of bundle compression.
– If (1) satisfies Assumption 11 and $\mathcal{Q}$ is finite, a variant of Algorithm 4 (with $m = 1$, $\hat{\mathcal{B}}_\ell = \emptyset$, and fixed prox-parameter at null steps) stops after a finite number of iterations with a primal solution of (1).

For these results to hold, the Separation Procedure must be "good enough" and able to eventually separate all violated inequalities.

When compared to an algorithm using a subgradient method for the dual step, the Dynamic Bundle Algorithm 4 has the important property of generating a descent (sub)sequence of iterates. This feature is crucial for proving convergence. So, at first sight, trying to extend our proofs to subgradient methods does not seem to be straightforward.

We assess our approach with two different examples: one where the Separation Procedure does satisfy the assumption (LOP), and another one where the assumption does not hold (TSP). Numerical results confirm the quality of the directions generated by the Dynamic Bundle Method. In the case of the TSP instances, due to the use of heuristic separation techniques, one might expect for the method to terminate before it can find all the right inequalities[4]. Finally, the main improvement effort must be focused on the QP performances as inequalities are added and removed along iterations: in average more than 50% of the total CPU time is spent by the QP solver. In its actual form, our code is suitable only for those cases where each Dual Evaluation is extremely expensive (for example when each subproblem involves the solution of a large-scale stochastic program, or the integration of large systems of nonlinear differential equations, or stochastic simulations).

References

[BA00] F. Barahona and R. Anbil. The volume algorithm: producing primal solutions with a subgradient method. *Math. Program.*, 87(3, Ser. A):385–399, 2000.

[Bea90] J. E. Beasley. A Lagrangian heuristic for set-covering problems. *Naval Res. Logist.*, 37(1):151–164, 1990.

[BGLS03] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Numerical Optimization. Theoretical and Practical Aspects.* Universitext. Springer-Verlag, Berlin, 2003. xiv+423 pp.

[BL02a] A. Belloni and A. Lucena. Improving on the Help and Karp Bound for the STSP via Lagrangian Relaxation. Working paper, 2002.

[BL02b] A. Belloni and A. Lucena. Lagrangian Heuristics to Linear Ordering. Working paper, 2002.

[4] An indicative of this effect is that the SG5K variant finds slightly better bounds.

[BMS02] L. Bahiense, N. Maculan, and C.A. Sagastizábal. The Volume Algorithm revisited. Relation with Bundle Methods. *Math. Program., Ser. A*, 94(1):41–69, 2002.

[CB83] N. Christofides and J. E. Beasley. Extensions to a Lagrangean relaxation approach for the capacitated warehouse location problem. *European J. Oper. Res.*, 12(1):19–28, 1983.

[CG59] E. Cheney and A. Goldstein. Newton's method for convex programming and tchebycheff approximations. *Numerische Mathematik*, 1:253–268, 1959.

[CLdS01] F. Calheiros, A. Lucena, and C. de Sousa. Optimal rectangular partitions. Technical report, Laboratório de Métodos Quantitativos, Departamento de Administracão, Universidade Federal do Rio de Janeiro, Brazil, 2001.

[EGM94] L. F. Escudero, M. Guignard, and K. Malik. A Lagrangian relax-and-cut approach for the sequential ordering problem with precedence relationships. *Ann. Oper. Res.*, 50:219–237, 1994. Applications of combinatorial optimization.

[Erm66] Ju. M. Ermol′ev. Method for solving nonlinear extremal problems. *Kibernetika (Kiev)*, 2(4):1–17, 1966.

[Fra02] A. Frangioni. Generalized bundle methods. *SIAM Journal on Optimization*, 13(1):117–156, 2002.

[GJR84] M. Grötschel, M. Jünger, and G. Reinelt. A cutting plane algorithm for the linear ordering problem. *Oper. Res.*, 32(6):1195–1220, 1984.

[GP79] C. Gonzaga and E. Polak. On constraint dropping schemes and optimality functions for a class of outer approximation algorithms. *SIAM Journal on Control and Optimization*, 17(4):477–497, 1979.

[GPT80] C. Gonzaga, E. Polak, and R. Trahan. An improved algorithm for optimization problems with functional inequality constraints. *IEEE Transactions on Automatic Control*, 25:49–54, 1980.

[Hel01] C. Helmberg. A cutting plane algorithm for large scale semidefinite relaxations. Technical Report ZIB-Report ZR-01-26, Konrad-Zuse-Zentrum Berlin, 2001. To appear in the Padberg Festschrift "The Sharpest Cut", SIAM.

[HFK01] M. Hunting, U. Faigle, and W. Kern. A Lagrangian relaxation approach to the edge-weighted clique problem. *European J. Oper. Res.*, 131(1):119–131, 2001.

[HUL93] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*. Number 305-306 in Grund. der math. Wiss. Springer-Verlag, 1993. (two volumes).

[HWC74] M. Held, Ph. Wolfe, and H.P. Crowder. Validation of subgradient optimization. *Math. Programming*, 6:62–88, 1974.

[Kel60] J. E. Kelley. The cutting plane method for solving convex programs. *J. Soc. Indust. Appl. Math.*, 8:703–712, 1960.

[Kiw85] K.C. Kiwiel. *Methods of descent for nondifferentiable optimization*. Springer-Verlag, Berlin, 1985.

[LB96] Abilio Lucena and John E. Beasley. Branch and cut algorithms. In *Advances in linear and integer programming*, volume 4 of *Oxford Lecture Ser. Math. Appl.*, pages 187–221. Oxford Univ. Press, New York, 1996.

[Luc92] A. Lucena. Steiner problem in graphs: Lagrangean relaxation and cutting-planes. *COAL Bulletin*, 21(2):2–8, 1992.

[MLM00] C. Martinhon, A. Lucena, and N. Maculan. A relax and cut algorithm for the vehicle routing problem. Technical report, Laboratório de Métodos Quantitativos, Departamento de Administracão, Universidade Federal do Rio de Janeiro, Brazil, 2000.

[Sho85] N. Shor. *Minimization methods for non-differentiable functions*. Springer-Verlag, Berlin, 1985.