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

A. Daniilidis · C. Lemaréchal

On a Primal-Proximal Heuristic in Discrete Optimization

the date of receipt and acceptance should be inserted later

Abstract. Lagrangian relaxation is useful to bound the optimal value of a given optimization problem, and also to obtain relaxed solutions. To obtain primal solutions, it is conceivable to use a convexification procedure suggested by D.P. Bertsekas in 1979, based on the proximal algorithm in the primal space.

The present paper studies the theory assessing the approach in the framework of combinatorial optimization. Our results indicate that very little can be expected in theory, even though fairly good practical results have been obtained for the unit-commitment problem.

Key words. Proximal algorithm, Lagrangian relaxation, primal-dual heuristics
Subject Classification (2000) 90C26, 49M27

1. Introduction, Motivation

This paper is motivated by a practical application: the unit-commitment problem, more precisely to optimize the generation schedules of the set of electrical power plants in France. Such a problem is usually solved through duality ([2, 12], see also [7,13] for additional references). After solving the dual comes the question of recovering a primal feasible solution, possibly suboptimal. An idea is to add in the production cost a quadratic term penalizing the deviation from the relaxed solution, obtained by dual means. This can give fairly good practical results, reported elsewhere: [5].

Our aim here is to study theoretically this approach, with emphasis on combinatorial problems lending themselves to Lagrangian relaxation. To this aim, we consider first the general optimization problem

$$\inf f(x)\,, \quad x \in \mathbb{R}^n\,. \tag{1.1}$$

In this simplified notation, possible constraints are incorporated into f via the indicator function (0 on the feasible set, $+\infty$ outside). We have particularly in mind discrete optimization, say

$$\min g(x)\,, \quad Ax \geqslant c \in \mathbb{R}^m\,, \quad x \in Z := \{x_1, \ldots, x_K\} \tag{1.2}$$

(where we could have a linear objective function $g(x) = b^\top x$); in this case f of (1.1) is

$$f(x) := \begin{cases} g(x) & \text{if } Ax \geqslant c \text{ and } x \in Z\,, \\ +\infty & \text{otherwise}\,. \end{cases} \tag{1.3}$$

INRIA, 655 avenue de l'Europe, Montbonnot 38334 Saint Ismier, France
e-mail: {aris.daniilidis,claude.lemarechal}@inrialpes.fr

Our aim is to study for such problems a convexification procedure introduced by D.P. Bertsekas in [1], based on the (primal) *proximal algorithm*.

1.1. The General Idea: Moreau Envelope

The basic idea of this procedure is to introduce the intermediate function

$$f_r(x,y) := f(x) + r\|x-y\|^2\,, \tag{1.4}$$

where $r > 0$ and $y \in \mathbb{R}^n$ is an additional variable, and to set

$$\varphi_r(y) := \inf\{f_r(x,y) : x \in \mathbb{R}^n\}\,. \tag{1.5}$$

Usually, φ_r is called the Moreau-Yosida regularization of f, originally assumed to be convex; for more general functions f see [16], where φ_r is called the Moreau envelope. Let us mention some intuitive facts, which will be stated more precisely in §1.3 below. Minimizing f (with respect to x) is "equivalent" to minimizing f_r (with respect to (x,y)), which in turn is "equivalent" to minimizing φ_r (with respect to y); and this amounts to finding y^* such that $f_r(\cdot,y^*)$ attains its minimum at the point $x = y^*$. In other words, minimizing f can be viewed as finding a fixed point of the so-called *proximal mapping* $y \mapsto \operatorname{Argmin} f_r(\cdot,y)$. This is the motivation for the proximal algorithm, which computes y_{k+1} by minimizing $f_r(\cdot,y_k)$.

Such a mechanism sounds highly artificial (why not set $r = 0$!). However observe that $f_r(\cdot,y)$ is "more convex" (for fixed y) than f and its minimization might therefore be easier. Now if φ_r is "sufficiently convex", then its minimization may be easy as well. Observe, incidentally, that the minimization of φ_r is an *unconstrained* problem (φ_r is $+\infty$ nowhere – unless $f \equiv +\infty$).

Keeping in mind the good results reported in [5], an interesting question is then whether the above approach (heuristic anyway) can be assessed by some theory. In a way, the present paper brings a negative answer to this last point. In the next subsection we give a flavor of our main results.

1.2. Content of the paper

Subsections §1.3 and §1.4 contain some general properties of φ_r and its minimum points, in the abstract case (1.1). Under mild assumptions, the local [global] minima of f coincide with the local [global] minima of φ_r. The proximal algorithm decreases the value $f(y_k)$ by a definite amount at each iteration (implying in particular that each y_k is feasible) and its cluster points are fixed points of the proximal mapping.

In the rest of the paper we focus our attention on (1.2), assuming first in §2 that φ_r can be computed exactly. Then the situation is not bad. Local minima of φ_r coincide with fixed points of the proximal mapping. The proximal mapping stops at a local minimum of φ_r, which is

– an optimal solution of (1.2) if r is small enough,
– or an arbitrary feasible point of (1.2) if r is large enough.

However, minimizing $f_r(\cdot, y)$ is normally as difficult as minimizing f. We therefore consider in §3 the case where φ_r has to be approximated via Lagrangian relaxation. Still considering (1.2), call

$$\tilde{\varphi}_r(y) := \max_{\lambda \geqslant 0} \min_{x \in Z} \left\{ g(x) + r\|x - y\|^2 - \lambda^\top (Ax - c) \right\} \tag{1.6}$$

the value obtained by minimizing $f_r(\cdot, y)$ through Lagrangian relaxation (note that $\tilde{\varphi}_r \leqslant \varphi_r$ because of the weak duality). The study of $\tilde{\varphi}_r$ involves substantial technicalities; here we give a superficial and informal description of our results.

Lagrangian relaxation is known to convexify Z in (1.2). Its effect on the (nonlinear) objective function of (1.6) is more complex; in fact it is possible to construct a function (which will be denoted by γ_r in §3.2) making it possible to express $\tilde{\varphi}_r$ as a min-function, just as φ_r. To understand what γ_r is, assume a linear problem: $g(x) = b^\top x$ in (1.2) and look at Fig. 1.1.

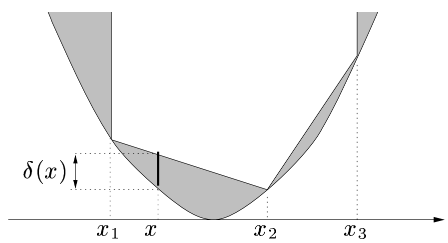

Fig. 1.1. Convexification of the squared norm over $Z := \{x_1, x_2, x_3\}$

The lower curve of Fig. 1.1 represents the squared norm $\|x\|^2$. The upper polygonal line represents the largest convex function whose value at each $x \in Z$ is lower than $\|x\|^2$ (in other words: the convex hull of the sum $\|x\|^2$ + indicator function of Z). Defining the function $\delta(x)$ to be the difference between the two, $g(x) = b^\top x$ in (1.6) can be replaced by $b^\top x + r\delta(x)$; in other words,

$$\tilde{\varphi}_r(y) = \min \left\{ b^\top x + r\delta(x) + r\|x - y\|^2 : x \in \operatorname{co} Z,\ Ax \geqslant c \right\}. \tag{1.7}$$

Remark 1.1. Because the squared norm is convex, $\delta(x) \geqslant 0$ for all x: the objective function of (1.7) is larger than in the original Lagrangian relaxation

$$\min \left\{ b^\top x : x \in \operatorname{co} Z,\ Ax \geqslant c \right\}. \tag{1.8}$$

Besides δ is 0 on Z – at least when g is linear. This has two consequences.

(i) The Moreau-Yosida regularization does reduce the duality gap, *even in the bad cases* where it merely reproduces a solution of (1.8); see a) below; such a solution will be denoted by x_c in the sequel. However, this property is of little use unless the *global* minimal value of $\tilde{\varphi}_r$ is known.

(ii) The feasible polyhedron in (1.8) has two kinds of extreme points: those in Z and "parasitic" ones, not lying in Z. The extra term $\delta(x)$ in (1.7) attracts the minimizers toward Z. Replacing φ_r by $\tilde{\varphi}_r$ is better than merely applying the Moreau-Yosida regularization to (1.8). □

Having thus put $\tilde{\varphi}_r$ in a form amenable to our framework, we can study the variant $y_{k+1} = \tilde{x}(y_k)$ of the proximal algorithm, where $\tilde{x}(y)$ is a primal solution of (1.6) or (1.7) (assumed to be computable). Note, however, that the Moreau-Yosida regularization is applied to the function $b^\top x + r\delta(x)$, which depends on r; this entails further technicalities. Note also that §2 does not apply since the feasible set in (1.7) is not finite.

The situation is now much less favourable than in §2. The sequence y_k has at least one cluster point, say y^*, which is a priori a local minimum of $\tilde{\varphi}_r$, and therefore of $b^\top x + r\delta(x)$ as well. Then several cases may occur:

a) If r is small, then y^* may be an optimal solution x_c of the relaxed problem (1.8).
b) If r is large, then y^* may be any feasible point in (1.2).
c) It may happen that $y^* \in Z$. Then we are roughly in the situation of §2: y^* is a local minimum of φ_r.

Finally, observe another deficiency of the implementable variant: nothing guarantees that the sequence $g(y_k)$ of objective values is monotone decreasing.

1.3. General Properties of the Moreau Envelope

Before specializing to discrete optimization problems, we give here a few general results relating the minimization of f and of φ_r.

Lemma 1.2. *The following general properties hold.*

(i) For all $y \in \mathbb{R}^n$, the function $r \mapsto \varphi_r(y)$ is nondecreasing.
(ii) $\varphi_r(y) \leqslant f(y)$ for all $y \in \mathbb{R}^n$ and all $r > 0$.

Assume that f is lower semicontinuous and bounded from below. Then:

(iii) if y_0 is a local minimum of φ_r then $x = y_0$ is the unique minimum point of $f_r(\cdot, y_0)$ in (1.5). In particular $\varphi_r(y_0) = f(y_0)$;
(iv) local minima of φ_r are also local minima of $\varphi_{r'}$ for $r' \geqslant r$.

Proof. (i) Obviously, $f_r(x, y)$ in (1.4) is a nondecreasing function of r and this property is transmitted to the infima.

(ii) Just observe that $\varphi_r(y) \leqslant f_r(y, y) = f(y)$.

(iii) Let $y = y_0$ in (1.5). Then our assumption implies that the infimum is attained at some x_0. Now for any y close to y_0, we can write

$$f(x_0) + r\|y_0 - x_0\|^2 = \varphi_r(y_0) \leqslant \varphi_r(y) \leqslant f(x_0) + r\|y - x_0\|^2 ,$$

hence $\|y_0 - x_0\|^2 \leqslant \|y - x_0\|^2$. Take in particular $y = y_0 - t(y_0 - x_0)$, so that $y - x_0 = (1-t)(y_0 - x_0)$. Then $\|y_0 - x_0\|^2 \leqslant (1-t)^2\|y_0 - x_0\|^2$. This is impossible if $\|y_0 - x_0\|^2 > 0$ and $t \in (0, 2)$, hence $\|y_0 - x_0\|^2 = 0$.

(iv) Using (i), we have for a local minimum y_0 of φ_r:

$$\varphi_{r'}(y) \geqslant \varphi_r(y) \geqslant \varphi_r(y_0) \quad \text{for } y \text{ close to } y_0 \;;$$

but from (iii) and (ii), $\varphi_r(y_0) = f(y_0) \geqslant \varphi_{r'}(y_0)$, which completes the proof. □

Remark 1.3. In these results, (iii) is the most important. An alternative proof can be given as follows. Being a min-function, φ_r is usually not differentiable: the concept of derivative, or gradient, is then replaced by that of *directional derivatives*:

$$\varphi_r'(y, d) := \lim_{t \downarrow 0} \frac{\varphi_r(y + td) - \varphi_r(y)}{t}, \qquad \text{for given } d \in \mathbb{R}^n \;.$$

Now a well-known formula (due to J.M. Danskin in [4]) says that, under appropriate assumptions on f_r, the directional derivative of functions given by (1.5) exists and has the expression

$$\varphi_r'(y, d) = \min \left\{ d^\top \nabla_y f_r(x, y) : \; x \text{ minimizes } f_r(\cdot, y) \right\}. \tag{1.9}$$

Here $\nabla_y f_r(x, y) = 2r(y - x)$ is the partial derivative of f_r with respect to y. In plain words: when moving from y to $y + td$ ($t > 0$ small), the marginal change of φ_r is the *smallest* scalar product of d with the partial derivatives of the minimand, computed at all the minimizing x's. For a local minimum, this change must be nonnegative: $\varphi_r'(y, d) \geqslant 0$ for any $d \in \mathbb{R}^n$, i.e. $d^\top \nabla_y f_r(x, y) \geqslant 0$ for any minimizing x and any $d \in \mathbb{R}^n$. This just means $\nabla_y f_r(x, y) = 2r(y - x) = 0$, i.e $x = y$ for any x minimizing (1.4). □

Lemma 1.2 (iii) suggests that points y such that $f_r(\cdot, y)$ is minimized at y (we use the notation $y \in \operatorname{Argmin} f_r(\cdot, y)$) play a special role. The following result says a little more about that:

Proposition 1.4. *If $y \in \operatorname{Argmin} f_r(\cdot, y)$, then y is the unique minimum of $f_{r'}(\cdot, y)$ for all $r' > r$.*

Proof. Just write that, for all $x \neq y$,

$$f_r(y, y) = f(y) \leqslant f(x) + r\|x - y\|^2 < f(x) + r'\|x - y\|^2$$

and observe that $f_{r'}(y, y) = f(y)$. □

Intuitively, minimizing φ_r in (1.5) is equivalent to minimizing f; this can be made precise:

Theorem 1.5. *The minimization of f and of φ_r are related as follows:*

(i) $\inf \left\{ f(x) : x \in \mathbb{R}^n \right\} = \inf \left\{ \varphi_r(y) : y \in \mathbb{R}^n \right\}$.

(ii) If x^* minimizes f, then x^* minimizes φ_r.

(iii) Assume f is lower semicontinuous and bounded from below. If y^* minimizes (resp. minimizes locally) φ_r then y^* minimizes (resp. minimizes locally) f.

Proof. (i) For any x and y in $\mathbb{R}^n$, $f_r(x,y) \geq f(x)$; hence $\varphi_r(y) \geq \inf f$ and $\inf \varphi_r \geq \inf f$. On the other hand, Lemma 1.2 (ii) gives $\inf \varphi_r \leqslant \inf f$.

(ii) In view of (i) and of Lemma 1.2 (ii), $\inf \varphi_r = \inf f = f(x^*) \geqslant \varphi_r(x^*)$.

(iii) It follows from Lemma 1.2 (ii), (iii) that $f(y^*) = \varphi_r(y^*) \leqslant \varphi_r(y) \leqslant f(y)$ for all $y \in \mathbb{R}^n$ (resp. for y close enough to y^*). □

Thus, the approach replaces a single minimization (of f) by the minimization of φ_r; this requires several computations of $\varphi_r(y)$, i.e. several minimizations of $f_r(\cdot, y)$, a function which is "more convex" than f. Then a natural question is whether φ_r has a chance of being convex; this holds at least when f itself is convex:

Theorem 1.6. *Assume f is bounded from below. Then*

$$f \text{ convex} \iff f_r \text{ convex (jointly)} \implies \varphi_r \text{ convex} .$$

Proof. For the second implication, the marginal function associated with a convex function is still convex; this is a classical result, see for example [10, Corollary B.2.4.5].

As for the first equivalence: if f is convex, then f_r is obviously convex (jointly with respect to x and y). If f_r is convex, then in particular $x \mapsto f(x) = f_r(x,x)$ is convex. □

In view of Theorem 1.5, better convexification properties could hardly be expected from this procedure. Indeed (1.1) contains just about any optimization problem; in particular, there are instances of (1.1) which are difficult, but for which f_r is convex in x and computing φ_r is "easy" – §3.6 will be devoted to a class of such problems. In these cases, minimizing f could not be equivalent to minimizing φ_r, if the latter were convex!

On the other hand, a classical convexification scheme is augmented Lagrangian [15]. Applied to a problem such as (1.2), for example, it would add the term $r\|Ax - c\|^2$. This also corresponds to introducing a Moreau envelope, but in the *dual* space; it does result in a convex optimization problem for r large enough but is hardly implementable; see [9, § XII.5.2] for example. As mentioned in [1], the present *primal* approach has the advantage of preserving separability of f, if any. The crucial point is that the quadratic term in (1.4) is a sum over the coordinates of the primal variable x; as such, it is not too complicating. In fact, we are interested in instances of (1.1) amenable to Lagrangian relaxation – such is the case of (1.2). Our aim will then be to reduce the duality gap, and/or to produce heuristic primal solutions, generated by the algorithm minimizing φ_r.

1.4. The Proximal Algorithm

The algorithm suggested in [1] to minimize φ_r is essentially

$$y_{k+1} \in \operatorname{Argmin} \{ f_r(x, y_k) : x \in \mathbb{R}^n \} , \tag{1.10}$$

where the set Argmin of global minimizers is assumed nonempty (this is the case when f is lower semicontinuous and bounded from below); naturally, the algorithm stops when $y_{k+1} = y_k$. This is called the *proximal algorithm*, whose convergence properties rely on the following result:

Theorem 1.7. *Assume* y_{k+1} *exists in* (1.10).

(i) There holds at each iteration:

$$f(y_{k+1}) \leqslant f(y_k) - r\|y_k - y_{k+1}\|^2 .$$

(ii) Assume that $f(y_k)$ *is bounded from below. Then* $\sum_k \|y_{k+1} - y_k\|^2 < +\infty$.

(iii) Assume further that f *is lower semicontinuous. Then any cluster point* y^* *of* y_k *is the unique minimum of* $f_{r'}(\cdot, y^*)$, *for any* $r' > r$.

Proof. By definition of y_{k+1} and from Lemma 1.2 (ii),

$$f(y_{k+1}) + r\|y_{k+1} - y_k\|^2 = \varphi_r(y_k) \leqslant f(y_k) ,$$

which is the stated inequality in (i).

We obtain by summation

$$f(y_{K+1}) - f(y_1) \leqslant -r \sum_{k=1}^{K} \|y_{k+1} - y_k\|^2 ,$$

which shows that the series $\sum_k \|y_{k+1} - y_k\|^2$ converges, unless $f(y_k) \downarrow -\infty$; (ii) follows.

For (iii), using again the definition of y_{k+1}, write for each k

$$f(x) + r\|x - y_k\|^2 \geqslant f(y_{k+1}) + r\|y_{k+1} - y_k\|^2 , \quad \text{for all } x \in \mathbb{R}^n .$$

Pass to the limit; f is lower semicontinuous and $\|y_{k+1} - y_k\| \to 0$ from (ii), so

$$f(x) + r\|x - y^*\|^2 = f_r(x, y^*) \geqslant f(y^*) = f_r(y^*, y^*) , \quad \text{for all } x \in \mathbb{R}^n ;$$

but this means $f_r(x, y^*) \geqslant f_r(y^*, y^*)$. The rest follows from Proposition 1.4. □

We explain the motivation of the proximal algorithm in the light of Remark 1.3. To avoid excessive generality, assume that φ_r is a smooth function, namely that it has a gradient $\nabla\varphi_r(y)$ at every $y \in \mathbb{R}^n$. Then its directional derivatives are $\varphi_r'(y, d) = d^\top \nabla\varphi_r(y)$ for all $d \in \mathbb{R}^n$; with Danskin's formula (1.9), this clearly implies that $f_r(\cdot, y)$ must have a *unique* minimizer[1] $x(y)$, and that $\nabla\varphi_r(y) = 2r[y - x(y)]$. In particular, each next iterate y_{k+1} in (1.10) exists and is defined without ambiguity. Besides, since

$$y_{k+1} = x(y_k) = y_k + \frac{1}{2r} 2r(x(y_k) - y_k) = y_k - \frac{1}{2r}\nabla\varphi_r(y_k) ,$$

we see that the proximal algorithm is just the minimization of φ_r by a standard gradient method. Now, from Theorem 1.7:

[1] To see this, observe that the directional derivative is linear in d, hence symmetric; with several minimizers, we would have $\varphi_r'(y, d) \neq -\varphi_r'(y, -d)$ for some direction d, a contradiction.

– either $f(y_k) \to -\infty$ (then we are certainly minimizing f successfully!)
– or $2r(y_k - y_{k+1}) = \nabla\varphi_r(y_k) \to 0$.
In the latter situation, assume that the sequence $\{y_k\}$ has some cluster point y^*. If φ_r is actually continuously differentiable (this means that the mapping $y \mapsto x(y)$ is *continuous*), then we see that $\nabla\varphi_r(y^*) = 2r[y^* - x(y^*)] = 0$: the proximal algorithm can only produce stationary points of φ_r, which have a chance to be local minimizers.

2. Discrete Optimization Problems: Conceptual Forms

In this section we focus on the particular case where (1.1) is actually an optimization problem on a *finite* set: we consider

$$\min g(x)\,, \quad x \in F = \{x_1, \dots, x_K\}\,. \tag{2.1}$$

The function g is left unspecified for the moment; it is completely characterized by its (finitely many) values $\{g(x_k)\}_{k=1}^K$. With respect to our previous notation,

$$f(x) = \begin{cases} g(x_k) & \text{if } x = x_k \text{ for some } k = 1, \dots, K\,, \\ +\infty & \text{otherwise}\,. \end{cases} \tag{2.2}$$

Needless to say, this function is bounded and lower semi-continuous: the assumptions made in §1 (especially Theorem 1.5) are trivially satisfied. As for the "outer" objective function, it becomes

$$\varphi_r(y) = \min_{x \in F} \left\{ g(x) + r\|x - y\|^2 \right\}. \tag{2.3}$$

First of all, local minima of φ_r can be characterized in this particular situation:

Proposition 2.1. *A point y_0 is a local minimum of φ_r if and only if $x = y_0$ is the unique optimal solution of (2.3) for $y = y_0$. In particular, every local minimum of φ_r is a strict local minimum.*

Proof. In view of Lemma 1.2 (iii), we have only to prove that, if $x = y_0$ is the unique solution of (2.3) for $y = y_0$, then y_0 is a strict local minimum of φ_r. Note first that, since F is a finite set, there is $\varepsilon > 0$ such that, for all $x \in F$ different from y_0,

$$g(x) + r\|x - y_0\|^2 \geqslant g(y_0) + 2\varepsilon = \varphi_r(y_0) + 2\varepsilon\,.$$

Now take y close enough to y_0 so that, for all $x \in F$,

$$r\|x - y\|^2 \geqslant r\|x - y_0\|^2 - \varepsilon\,.$$

Summing these two inequalities, we obtain

$$g(x) + r\|x - y\|^2 \geqslant \varphi_r(y_0) + \varepsilon$$

and hence

$$\min\left\{g(x)+r\|x-y\|^2 : x\in F\backslash\{y_0\}\right\}\geqslant\varphi_r(y_0)+\varepsilon\,.$$

As a result: for y close enough to y_0,

$$\varphi_r(y)\geqslant\min\left\{\varphi_r(y_0)+\varepsilon,\varphi_r(y_0)+r\|y_0-y\|^2\right\}\geqslant\varphi_r(y_0)\,,$$

and the second inequality is strict if $y\neq y_0$. □

Note that φ_r is the minimum of finitely many quadratic functions. The y-space is divided into regions inside which the minimum in (2.3) is attained at some point $x\in F$, call it x_k (k depending on the region in question). The corresponding quadratic portion has the equation $g(x_k)+r\|y-x_k\|^2$, with gradient $2r(y-x_k)$. Altogether, φ_r looks as indicated in Fig. 2.1. Local minima can be points such as y_1 or y_2; but at y_3, there are two minimum points in (2.3): $x=y_3$ and some other $x\in F$; in view of Lemma 1.2 (iii), y_3 cannot be a local minimum of φ_r.

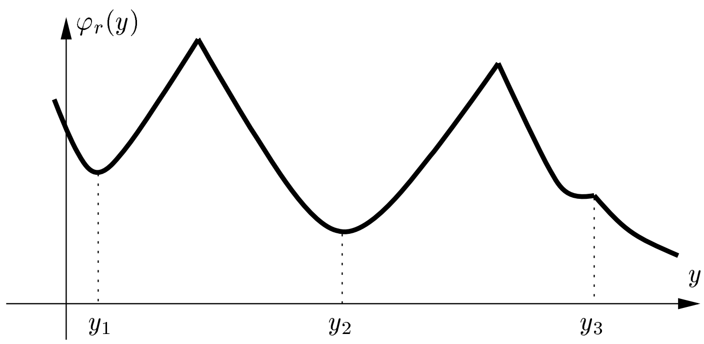

Fig. 2.1. A piecewise quadratic function

We already know from Proposition 2.1 that local minima of φ_r lie in F. Our next result specifies which feasible points can be thus obtained.

Theorem 2.2. *For any $r>0$, any local minimum of φ_r lies in F. Moreover:*

(i) For r large enough, the local minima of φ_r are exactly the points in F.

(ii) For r small enough, the local minima of φ_r are exactly the optimal solutions of (2.1)*.*

Proof. If y_0 is a local minimum of φ_r, Proposition 2.1 shows that y_0 minimizes $f_r(x,y_0)=g(x)+r\|x-y_0\|^2$ over $x\in F$; then in particular $y_0\in F$.

(i) We have to prove that an arbitrary $y_0\in F$ is a local minimum of φ_r if r is large enough. Define the diameter of $g(F)$

$$\Gamma:=\max\left\{g(x)-g(x'):x\in F,x'\in F\right\},$$

the "discreteness" of F

$$\varepsilon := \min\{\|x - x'\| : x \in F, x' \in F, x \neq x'\},$$

and take $r > \Gamma/\varepsilon^2$.

For y close to y_0, namely for $\|y - y_0\| \leqslant \delta := \varepsilon - \sqrt{\Gamma/r} > 0$, there holds for all $x \in F$ different from y_0:

$$\|x - y\| \geqslant \|x - y_0\| - \|y_0 - y\| \geqslant \varepsilon - \delta = \sqrt{\Gamma/r}.$$

Then we write

$$\begin{aligned} g(x) + r\|y - x\|^2 &\geqslant g(y_0) + g(x) - g(y_0) + r(\varepsilon - \delta)^2 \\ &\geqslant g(y_0) - \Gamma + r(\varepsilon - \delta)^2 \\ &= g(y_0) \geqslant \varphi_r(y_0), \end{aligned}$$

where the last inequality is Lemma 1.2 (ii). The conclusion follows by taking the infimum over x (knowing that the inequality holds trivially for $x = y_0$).

(ii) Call v^* the optimal value in (2.1) and V^* the set of optimal solutions; let

$$v^+ := \min\{g(x) : x \in F \backslash V^*\}$$

the "next to optimal" value of g over F, and finally let

$$D := \max\{\|x - x'\| : x, x' \in F\}$$

the diameter of F ($D > 0$ – except in the trivial case where F is a singleton). Note that $v^+ > v^*$ and take $0 < r < (v^+ - v^*)/D^2$.

Take $x^* \in V^*$ (so that $g(x^*) = v^+$) and let y be a local minimizer of φ_r. We already know that $y \in F$ and that $\varphi_r(y) = g(y)$ (Proposition 2.1); then write

$$g(y) = \varphi_r(y) \leqslant g(x^*) + r\|x^* - y\|^2 \leqslant g(x^*) + rD^2 < g(x^*) + v^+ - v^* = v^+.$$

By definition of v^+, $g(y)$ has to be equal to v^*; from Theorem 1.5 (i), this means $y \in V^*$. □

The proximal algorithm of §1.4 can then be described in the present context:

Algorithm 2.3 (Conceptual proximal algorithm) *Choose* $r > 0$.

Step 0. *Take* y_1 *arbitrary in* $\mathbb{R}^n$; *set* $k = 1$.

Step 1. *Let* x_k *realize the smallest of the numbers* $g(x_h) + r\|y_k - x_h\|^2$, *for* $h = 1, \dots, K$.

Step 2. *If* $x_k = y_k$ *then stop.*

Step 3. *Set* $y_{k+1} = x_k$, *replace* k *by* $k + 1$ *and go to Step 1.* □

Convergence is easy to establish:

Theorem 2.4. *The above algorithm terminates at some* k *with a* $y_k \in F$ *which is a local minimum of* $\varphi_{r'}$ *for any* $r' > r$.

Proof. The y_k's can take on finitely many values; but in view of Theorem 1.7, $y_{k+1} - y_k$ tends to 0; hence $y_{k+1} - y_k = 0$ at some k. By construction, this means $x_k = y_k$ minimizes the function $x \mapsto g(x) + r\|x - y_k\|^2$:

$$\varphi_r(y_k) = g(y_k) \leqslant g(x) + r\|x - y_k\|^2 , \quad \text{for all } x \in F ;$$

use Proposition 1.4 and Proposition 2.1 to finish the proof. □

Let us sum up this §2.

(i) First remember Lemma 1.2 (iv): when r grows from 0 to $+\infty$, the local minima of φ_r form nested sets, growing from the "ideal" set of optimal solutions of (2.1), to the "worst" whole feasible set. It is therefore advantageous to take a "small" r (whatever this means).

(ii) The proximal algorithm produces such a local minimum – a feasible point for (2.1). In terms of the objective function g, the quality of this point depends on
- the initialization: in view of Theorem 1.7, the objective function is improved by at least $r\|y_{k+1} - y_k\|^2$ at each iteration;
- the value of r: in view of Theorem 2.2, only an optimal solution can be produced if r is small enough.

(iii) If we were able to guarantee a *global* minimum of φ_r, instead of local, then we would for sure have an optimal solution, no matter how r was chosen (Theorem 1.5).

Unfortunately, this algorithm is only conceptual anyway: φ_r cannot be computed exactly – a fortiori minimized globally.

3. Discrete Optimization Problems: Implementable Forms

From now on, we assume the feasible set F of (2.1) to be some *structured* finite set amenable to Lagrangian relaxation: our problem becomes

$$\min g(x) , \quad x \in F := \{x \in Z : Ax \geqslant c \in \mathbb{R}^m\} , \quad Z \text{ finite in } \mathbb{R}^n . \tag{3.1}$$

Common instances in combinatorial optimization have

$$Z = \left\{x \in \{0,1\}^n : Bx \geqslant d \in \mathbb{R}^p\right\} . \tag{3.2}$$

The important feature is that (3.1) becomes "easy" when its linear constraints are relaxed. The "proximized" version becomes

$$\varphi_r(y) = \min \{g(x) + r\|x - y\|^2 : x \in Z, Ax \geqslant c\} , \tag{3.3}$$

also assumed easy if $Ax \geqslant c$ is relaxed.

3.1. Introducing Lagrangian Relaxation

Here, with the extra variable $\lambda \in \mathbb{R}^m_+$, we introduce the Lagrangian associated with (3.3) and the corresponding relaxation $\tilde{\varphi}_r$ of φ_r:

$$\begin{aligned} \tilde{\varphi}_r(y) &:= \sup_{\lambda \geqslant 0} \theta_r(\lambda, y)\,, \qquad \text{where} \\ \theta_r(\lambda, y) &:= \min_{x \in Z} \left\{ g(x) + \lambda^\top (c - Ax) + r\|x - y\|^2 \right\}. \end{aligned} \tag{3.4}$$

We will denote respectively by

$$v^* := \min_{x \in F} g(x) \quad \text{and} \quad v_c := \sup_{\lambda \geqslant 0} \min_{x \in Z} \left\{ g(x) + \lambda^\top (c - Ax) \right\} \tag{3.5}$$

the optimal value of (3.1) and of its relaxation (naturally $v^* = \varphi_0(y)$ and $v_c = \tilde{\varphi}_0(y)$ for all y).

Lemma 3.1. *The function $r \mapsto \tilde{\varphi}_r(y)$ is nondecreasing; $v_c \leqslant \tilde{\varphi}_r(y) \leqslant \varphi_r(y)$ for all y. It follows that*

$$v_c \leqslant \inf_{y \in \mathbb{R}^n} \tilde{\varphi}_r(y) \leqslant v^*\,.$$

Proof. Weak duality (see [11] for example) guarantees $\tilde{\varphi}_r(y) \leqslant \varphi_r(y)$. Furthermore, the Lagrangian $g(x) + \lambda^\top (c - Ax) + r\|x - y\|^2$ in (3.4) is obviously a nondecreasing function of r; and this property is transmitted to the infima and suprema. The rest follows easily, use in particular (3.5) and Theorem 1.5 (i). □

Thus, the "proximized" dualization (3.4) can potentially improve the duality gap (if one is able to minimize $\tilde{\varphi}_r$); Proposition 3.9 will tell more.

To express $\tilde{\varphi}_r$ in a form more amenable to calculations than (3.4), we lift the formulation (3.1) in the graph-space: we define the set

$$Z'_r := \left\{ (x, v) : x \in Z, v \geqslant g(x) + r\|x\|^2 \right\} \subset \mathbb{R}^n \times \mathbb{R}\,,$$

which is nothing but the epigraph[2] $\operatorname{epi} g_r$ of the function g_r defined by

$$\mathbb{R}^n \ni x \mapsto g_r(x) := \begin{cases} g(x) + r\|x\|^2 & \text{if } x \in Z, \\ +\infty & \text{otherwise}\,. \end{cases} \tag{3.6}$$

Developing $\|x - y\|^2$, we can then formulate (3.3) as

$$\begin{aligned} \varphi_r(y) &= r\|y\|^2 + \min_x \left\{ g_r(x) - 2ry^\top x : x \in \mathbb{R}^n, Ax \geqslant c \right\} \\ &= r\|y\|^2 + \min_{(x,v)} \left\{ v - 2ry^\top x : (x, v) \in Z'_r, Ax \geqslant c \right\}. \end{aligned}$$

The second line above reveals a linear Lagrangian: indeed θ_r in (3.4) can also be written

$$\theta_r(\lambda, y) := r\|y\|^2 + \min_{(x,v) \in Z'_r} \left[v - 2ry^\top x + \lambda^\top (c - Ax) \right] \tag{3.7}$$

and the well-known convexification effect of Lagrangian relaxation can be invoked. Convexifying Z'_r is a rather simple operation because it is confined to the convex hull of Z, a *compact* set.

[2] Recall that the epigraph of a function f is $\operatorname{epi} f := \{(x, v) : v \geqslant f(x)\}$. It will be useful in the sequel to keep in mind that, as a general rule, the symbol $'$ (as in Z') will denote a lifting into the graph-space $\mathbb{R}^n \times \mathbb{R}$.

Theorem 3.2. *The function $\tilde\varphi_r$ has the equivalent expression*

$$\tilde\varphi_r(y) = r\|y\|^2 + \min_{(x,v)} \left\{ v - 2ry^\top x : (x,v) \in \operatorname{co} Z'_r, Ax \geqslant c \right\} . \tag{3.8}$$

Proof. Using the form (3.7), this is a classical result; see for example [6, Thm 3.5], [8, Thm 1(d)], [14, Lemma 2.2], [13, §2.3] or [11, §3.2] – among others. □

The introduction of (3.8) is useful to express $\tilde\varphi_r$ as a min function, in contrast with (3.4). However the proximal algorithm of the previous sections operates in the x-space, making it necessary to eliminate the v-variable from (3.8). This introduces some technicalities, dealt with in our next subsection.

3.2. Some Technical Results

Recall in our context that the convex hull (biconjugate) f^{**} of a function f of the form (2.2) can equivalently be defined as

- the function whose epigraph is the convex hull of the epigraph of f: notationally, $\operatorname{epi} f^{**} = \operatorname{co} \operatorname{epi} f$;
- or the largest convex function smaller than f,
- or $f^{**}(x) := \min \sum_k \alpha_k f(x_k)$, where the sum runs over all sets of points x_k and convex multipliers α_k such that $\sum_k \alpha_k x_k = x$.

See for example [10, § B.2.5], more particularly Example B.2.5.5; see also Fig. 3.1.

Remark 3.3. In our context, given an arbitrary $y \in \operatorname{co} Z$, we can say that

- for any convex combination $\sum_k \alpha_k x_k$ of points x_k in Z with $\sum_k \alpha_k x_k = y$, there holds $g_r^{**}(y) \leqslant \sum_k \alpha_k g_r(x_k)$ (for all r, including $r = 0$);
- there is a convex combination $\sum_k \alpha_k x_k$ of points x_k in Z with $\sum_k \alpha_k x_k = y$, and such that $g_r^{**}(y) = \sum_k \alpha_k g_r(x_k)$. □

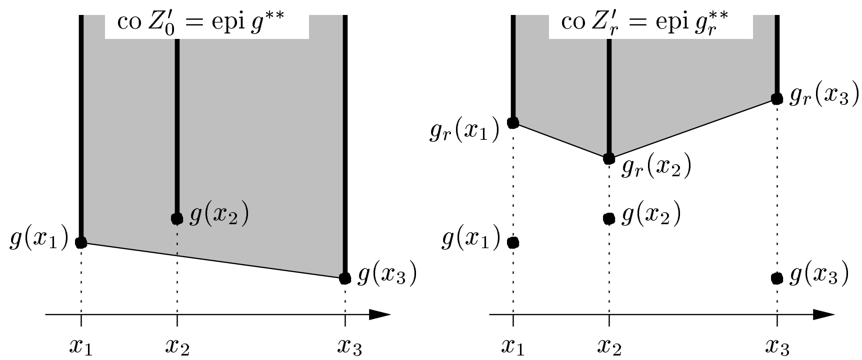

Fig. 3.1. Convexifying $\operatorname{epi} g$ as $\operatorname{epi} g_r$, with $Z := \{x_1, x_2, x_3\}$

Accordingly, the convex hull of Z'_r, introduced in (3.8), is itself the epigraph $\operatorname{epi} g_r^{**}$ of the biconjugate of the function g_r defined in (3.6). It is useful to understand that Z'_r is made up of finitely many vertical half-lines: one for each $x \in Z$, with bottom at altitude $g_r(x) = g(x) + r\|x\|^2$. Its convex hull $\operatorname{co} Z'_r$ is a convex polyhedron, unbounded from above, whose x-part is the bounded polyhedron $\operatorname{co} Z$, and having finitely many extreme points. Alternatively, the convex hull g_r^{**} of g_r is a polyhedral function with domain $\operatorname{co} Z$. The x-part of any extreme point of $\operatorname{co} Z'_r = \operatorname{epi} g_r^{**}$ certainly lies in Z; but conversely, a point in Z need not be the projection of an extreme point of $\operatorname{epi} g_r^{**}$ (see for example the point x_2 in the left part of Fig. 3.1).

Now the feasible set in (3.8) can be denoted by

$$\hat{Z}'_r := \{(x,v) \in \operatorname{co} Z'_r : Ax \geqslant c\},$$

and its extreme points form an important object since the minimand in (3.8) is linear. Denoting by $\operatorname{ext} F$ the set of extreme points of a set F, we define

$$\hat{F}'_r := \operatorname{ext}\{(x,v) \in \operatorname{co} Z'_r : Ax \geqslant c\},$$

which we project onto the x-space:

$$\hat{F}_r := \{x \in \operatorname{co} Z : (x,v) \in \hat{F}'_r \text{ for some } v\}. \tag{3.9}$$

Lemma 3.4. *Define the function* $\gamma_r(x) := g_r^{**}(x) - r\|x\|^2$. *Then*

$$\begin{aligned}
\tilde{\varphi}_r(y) &= r\|y\|^2 + \min\nolimits_x \{g_r^{**}(x) - 2ry^\top x : x \in \hat{F}_r\} && (a)\\
&= \min\nolimits_x \{\gamma_r(x) + r\|y-x\|^2 : x \in \hat{F}_r\} && (b)\\
&= \min\nolimits_x \{\gamma_r(x) + r\|y-x\|^2 : x \in \operatorname{co} Z, Ax \geqslant c\}. && (c)
\end{aligned} \tag{3.10}$$

Proof. Consider the linear program (3.8).

(i) The x-part [resp. v-part] of its feasible set is bounded [resp. from below]: the minimum value is finite, and is therefore attained at some extreme point:

$$\tilde{\varphi}_r(y) = r\|y\|^2 + \min_{(x,v)} \{v - 2ry^\top x : (x,v) \in \hat{F}'_r\}.$$

(ii) In fact, only the lower bound of v for each x need be considered; but this lower bound is $g_r^{**}(x)$:

$$\tilde{\varphi}_r(y) = r\|y\|^2 + \min_x \{g_r^{**}(x) - 2ry^\top x : x \in \hat{F}_r\},$$

which is (3.10)(a). Rearrange the terms to obtain (b). Finally, (c) is obtained by performing (ii) (and rearranging the terms) directly on (3.8). □

It is worth mentioning that the solutions of (3.10) for $r = 0$ solve the convexified primal problem: such solutions x_c produce the relaxed value $v_c = g^{**}(x_c)$. The following result confirms the intuitive property that *all* the useful values of g_r are preserved by g_r^{**} if r is large enough (see the right part of Fig. 3.1); then $\hat{F}_r$ contains the whole feasible set F of (3.1) – plus some parasitic points due to the intersection with the constraint $Ax \geqslant c$.

Lemma 3.5. *There exists* $\bar{r}_g \geqslant 0$ *such that, whenever* $r > \bar{r}_g$*:*

(i) $g_r^{**}(x) = g_r(x)$, *for all* $x \in Z$,

(ii) the extreme points of $\operatorname{co} Z'_r = \operatorname{epi} g_r^{**}$ *are exactly the points* $(x, g_r(x))$*, with* x *describing* Z*,*

(iii) $F \subset \hat{F}_r$*.*

These properties hold with $\bar{r}_g = 0$ *if* g *is the restriction to* Z *of a convex function.*

Proof. Interpolate g by some polynomial p (that is, $p = g$ on Z). Let λ be a lower bound on the minimal eigenvalue of the Hessian $\nabla^2 p(x)$ over $x \in \operatorname{co} Z$ (a bounded set!). If $2r + \lambda > 0$, i.e. if $r > \bar{r}_g := -\lambda/2$, then the function $p_r := p + r\|\cdot\|^2$ is strictly convex over $\operatorname{co} Z$.
(i) By construction, $p_r(x) = g_r(x)$ for any $x \in Z$; besides, $p_r(x) \leqslant g_r(x) = +\infty$ for $x \notin Z$. Altogether, p_r is a convex function below g_r, hence $p_r \leqslant g_r^{**} \leqslant g_r$. In particular,

$$p_r(x) = g_r(x) \leqslant g_r^{**}(x) \leqslant g_r(x) \quad \text{if } x \in Z \;,$$

and this is actually a chain of equalities.
(ii) Now any extreme point (x, v) of $\operatorname{co} Z'_r$ must
– satisfy $v = g_r(x)$ (otherwise, with $\varepsilon > 0$ so small that $v - \varepsilon \geqslant g_r(x)$ we would have the contradiction $(x, v) = \frac{1}{2}(x, v - \varepsilon) + \frac{1}{2}(x, v + \varepsilon)$) and
– lie in Z'_r (convexification cannot create extreme points).
Conversely, let $(x, g_r(x))$ lie in Z'_r. Then $g_r(x) = p_r(x)$ (since $x \in Z$), and strict convexity of p_r implies that $(x, g_r(x)) = (x, p_r(x))$ is extreme[3] in $\operatorname{epi} p_r$. *A fortiori*, it is extreme in $\operatorname{co} Z'_r = \operatorname{epi} g_r^{**} \subset \operatorname{epi} p_r$.
(iii) In the stated situation, any point $x \in Z$ satisfying $Ax \geqslant c$ (i.e. any point in F) certainly lies in $\hat{F}_r$.

Finally, when g has a convex extension, p can be taken as that extension, instead of a polynomial; then p_r is strictly convex for any $r > 0$ and the above arguments still hold. □

The set $\operatorname{co} Z'_r$ was usefully viewed as an epigraph. Now it is just as useful to see the feasible set of (3.8) as an epigraph:

$$\hat{Z}'_r := \operatorname{co} Z'_r \cap \{(x, v) : Ax \geqslant c\} = \operatorname{epi} \hat{g}_r^{**} \,,$$

where the function $\hat{g}_r^{**}$ is defined by

$$\hat{g}_r^{**}(x) := \begin{cases} g_r^{**}(x) & \text{if } Ax \geqslant c \;, \\ +\infty & \text{otherwise} \;. \end{cases}$$

Naturally, Remark 3.3 also applies to $\hat{g}_r^{**}$ whenever the considered point $y \in Z$ satisfies $Ay \geqslant c$.

Lemma 3.6. *The set* $\hat{F}_\infty := \cup_{r \geqslant 0} \hat{F}_r$ *is finite. In particular,*

$$d := \min \{\|x - y\|^2 : x, y \in \hat{F}'_\infty, x \neq y\} > 0 \,.$$

[3] For a strictly convex function f, all points $(x, f(x))$ are extreme in $\operatorname{epi} f$.

Proof. Call x_k the points of Z: $Z = \{x_1, x_2, \dots, x_K\}$ and introduce a "canonical epigraph"

$$\hat{\Gamma}' := \Big\{(\alpha, t) \in \mathbb{R}^K \times \mathbb{R} : \sum_k \alpha_k = 1, \sum_k \alpha_k A x_k \geqslant c, (\alpha, t) \geqslant 0\Big\}.$$

We claim that $\operatorname{epi} \hat{g}_r^{**}$ is the image of $\hat{\Gamma}'$ under the linear operator T_r which, to $(\alpha, t) \in \mathbb{R}^K \times \mathbb{R}$, associates

$$T_r(\alpha, t) := \begin{pmatrix} x(\alpha, t) \\ v_r(\alpha, t) \end{pmatrix} := \begin{pmatrix} \sum_k \alpha_k x_k \\ \sum_k \alpha_k g_r(x_k) + t \end{pmatrix}.$$

Indeed, if $(\alpha, t) \in \hat{\Gamma}'$, then $x(\alpha, t) \in \operatorname{co} Z$ and $T_r(\alpha, t) \in \operatorname{epi} g_r^{**}$; but we even have $T_r(\alpha, t) \in \operatorname{epi} \hat{g}_r^{**}$ since $Ax(\alpha, t) \geqslant c$. Conversely, let (x, v) with $v \geqslant \hat{g}_r^{**}(x)$ and take convex multipliers α_k realizing the minimal value of $\sum_k \alpha_k g_k(x_k)$, subject to $\sum_k \alpha_k x_k = x$: this value is just $\hat{g}_r^{**}(x)$ (Remark 3.3). Hence

$$t := v - \sum_k \alpha_k g_k(x_k) = v - \hat{g}_r^{**}(x) \geqslant 0$$

and $(\alpha, t) \in \hat{\Gamma}'$.

Now let $x \in \hat{F}_r$ for some $r \geqslant 0$ and let v_r be such that (x, v_r) is extreme in $\operatorname{epi} \hat{g}_r^{**}$. The pre-image by T_r of (x, v_r) is a face of the convex polyhedron $\hat{\Gamma}'$; every (α, t) in this face satisfies $\sum_k \alpha_k x_k = x$. Because the set $\hat{\Gamma}'$ (which does not depend on r) has finitely many faces, $\hat{F}_\infty$ can only be finite. □

3.3. Local Minima of $\tilde{\varphi}_r$

First of all, Proposition 2.1 can be reproduced as such:

Proposition 3.7. *A point y_0 is a local minimum of $\tilde{\varphi}_r$ if and only if $x = y_0$ is the unique optimal solution of* (3.10) *for $y = y_0$, or equivalently $(x, v) = (y_0, g_r^{**}(y_0))$ is the unique optimal solution of* (3.8)*.*

Proof. The fact that γ_r in (3.10) depends on r has no impact on Proposition 2.1. As for the second form, just apply the definitions. □

Lemma 3.5 has revealed an $\bar{r}_g \geqslant 0$, whose role is to compensate possible nonconvexities present in g; this key property will appear continually in what follows. We begin by mimicking Proposition 1.4.

Proposition 3.8. *Let $r \geqslant \bar{r}_g$ of Lemma 3.5 and suppose that* (3.10) *has an optimal solution $x = y$ which lies in Z (i.e. is feasible in* (3.1)*). Then this y is a local minimum of $\tilde{\varphi}_{r'}$ for all $r' > r$.*

Proof. In view of Lemma 3.5 (i), $\gamma_r(y) = g_r(y) - r\|y\|^2 = g(y)$. Then the proof goes as in Proposition 1.4. □

With respect to the pure Lagrangian relaxation, the proximal term does improve the duality gap strictly.

Proposition 3.9. *Take $r > \bar r_g$ of Lemma 3.5. If there is a duality gap (i.e. $v_c < v^*$), then any local minimizer y_0 of $\tilde\varphi_r$ satisfies $\tilde\varphi_r(y_0) > v_c$.*

Proof. With r as stated, let y_0 be a local minimum of $\tilde\varphi_r$. In view of Proposition 3.7, $x = y_0$ is the optimal solution in (3.10)(b) with $y = y_0$, hence

$$\tilde\varphi_r(y_0) = \gamma_r(y_0) = g_r^{**}(y_0) - r\|y_0\|^2 \,. \tag{3.11}$$

Suppose first $y_0 \in Z$. Then (Lemma 3.5 (i)) $g_r^{**}(y_0) = g_r(y_0)$; thus, from the definition (3.6) of g_r,

$$\tilde\varphi_r(y_0) = g(y_0) + r\|y_0\|^2 - r\|y_0\|^2 = g(y_0) \geqslant v^*$$

because y_0 is indeed feasible in (3.1).

If $y_0 \notin Z$ then $(y_0, g_r^{**}(y_0))$ is not extreme in $\operatorname{co} Z'_r$ (convexification cannot create extreme points): there is a nontrivial convex combination of points $(x_k, v_k) \in \operatorname{epi} g_r$ (i.e. $x_k \in Z$ and $v_k \geqslant g_r(x_k)$) such that

$$y_0 = \sum_k \alpha_k x_k \,, \quad g_r^{**}(y_0) = \sum_k \alpha_k v_k \geqslant \sum_k \alpha_k g_r(x_k) \,.$$

Combining with (3.11), we deduce

$$\begin{aligned}
\tilde\varphi_r(y_0) &\geqslant \textstyle\sum_k \alpha_k g_r(x_k) - r\|y_0\|^2 && \\
&= \textstyle\sum_k \alpha_k g(x_k) + r\sum_k \alpha_k\|x_k\|^2 - r\|y_0\|^2 && \text{[definition (3.6) of } g_r\text{]}\\
&> \textstyle\sum_k \alpha_k g(x_k) + r\|y_0\|^2 - r\|y_0\|^2 && \text{[strict convexity of } \|\cdot\|^2\text{]}\\
&\geqslant g^{**}(y_0) && \text{[Remark 3.3]}\\
&\geqslant \tilde\varphi_0(y_0) && [y_0 \text{ is feasible in (3.10)]}
\end{aligned}$$

and the proof is complete since we know that $\tilde\varphi_0(y_0) = v_c$. □

This result echoes Lemma 3.1: providing that g_r looks convex, the duality gap is strictly improved – although computing this duality gap implies as before the *global* minimization of $\tilde\varphi_r$. Now we turn to echoing Theorem 2.2.

Proposition 3.10. *For r large enough, the set of local minima of $\tilde\varphi_r$ contains the whole feasible set F of* (3.1).

Proof. Use the notation $\max_Z g$ [resp. $\min_Z g$] for the max [resp. the min] of g over Z. With d of Lemma 3.6, take $r > \max\{\bar r_g, (\max_Z g - \min_Z g)/d\}$, $y \in F$ and $x \in \hat F_r$ different from y. First of all, y is feasible in (3.10)(a) (recall from Lemma 3.5 (iii) that $F \subset \hat F_r$). Next, $g_r^{**}(x) = \sum_k \alpha_k g_r(x_k)$, for some convex combination of points in Z making up x (Remark 3.3). Then we apply the definition of g_r:

$$\begin{aligned}
g_r^{**}(x) &= \textstyle\sum_k \alpha_k g(x_k) + r\sum_k \alpha_k\|x_k\|^2 \\
&\geqslant \textstyle\sum_k \alpha_k g(x_k) + r\|x\|^2 && [\text{convexity of } \|\cdot\|^2] \\
&\geqslant \min_Z g + r\|x\|^2 && [\text{each } x_k \text{ lies in } Z] \\
&\geqslant \min_Z g + g(y) - \max_Z g + r\|x\|^2 && [y \in F \subset Z] \\
&> g(y) - rd + r\|x\|^2 && [\text{choice of } r] \\
&\geqslant g(y) - r\|x-y\|^2 + r\|x\|^2 && [x \neq y \in F \subset \hat{F}_\infty] \\
&= g(y) + 2ry^\top x - r\|y\|^2 .
\end{aligned}$$

Thus we have proved

$$g_r^{**}(x) - 2ry^\top x > g(y) - r\|y\|^2 = g_r(y) - 2r\|y\|^2 \geqslant g_r^{**}(y) - 2r\|y\|^2 .$$

In a word, y is the unique minimum point in (3.10)(a); as such, it minimizes $\tilde{\varphi}_r$ locally (Proposition 3.7). □

Let us now study global minima of $\tilde{\varphi}_r$. First, the upper bound v^* is attained in Lemma 3.1.

Theorem 3.11. *For r large enough, the global minima of $\tilde{\varphi}_r$ are exactly the optimal solutions of* (3.1).

Proof. For $y \notin Z$, let $\varepsilon(y)$ be the smallest value of $\sum_k \alpha_k\|x_k\|^2 - \|y\|^2$ over all convex combinations of points $x_k \in Z$ such that $\sum_k \alpha_k x_k = y$; strict convexity of the squared norm implies $\varepsilon(y) > 0$ and, since $\hat{F}_\infty \backslash Z$ is a finite set,

$$\varepsilon := \min\{\varepsilon(y) : y \in \hat{F}_\infty \backslash Z\} > 0 .$$

Then take $r > \max\{\bar{r}_g, (v^* - \min_Z g)/\varepsilon\}$ and let y_0 be a local minimum of $\tilde{\varphi}_r$.

Suppose for contradiction that $y_0 \notin F$; because y_0 is feasible in (3.10)(c), the definition of F in (3.1) shows that $y_0 \notin Z$. Then write from (3.10)(a)

$$\begin{aligned}
\tilde{\varphi}_r(y_0) &= g_r^{**}(y_0) - r\|y_0\|^2 && [\text{Proposition 3.7}] \\
&= \textstyle\sum_k \alpha_k g_r(x_k) - r\|y_0\|^2 && [\text{Remark 3.3}] \\
&= \textstyle\sum_k \alpha_k g(x_k) + r\big(\sum_k \alpha_k\|x_k\|^2 - \|y_0\|^2\big) && [\text{definition (3.6) of } g_r] \\
&\geqslant \min_Z g + r\varepsilon . && [\text{definition of } \varepsilon]
\end{aligned}$$

Thanks to the choice of r, $\tilde{\varphi}_r(y_0) > v^*$, i.e. y_0 cannot minimize $\tilde{\varphi}_r$ globally (Lemma 3.1). As a result, every global minimum of $\tilde{\varphi}_r$ certainly lies in F.

Let y_0 minimize $\tilde{\varphi}_r$ globally and use Lemma 3.5 (i): because $y_0 \in F \subset Z$,

$$\tilde{\varphi}_r(y_0) = g_r^{**}(y_0) - r\|y_0\|^2 = g_r(y_0) - r\|y_0\|^2 = g(y_0) \geqslant v^* .$$

In view of Lemma 3.1, y_0 solves (3.1): in fact, $\inf_y \tilde{\varphi}_r(y) = \tilde{\varphi}_r(y_0) = v^*$.

Conversely, let y^* solve (3.1):

$$g(y^*) = v^* = \inf \tilde{\varphi}_r \leqslant \tilde{\varphi}_r(y^*) \leqslant \varphi_r(y^*) \leqslant g(y^*) ,$$

where we have used successively Lemma 3.1 and Lemma 1.2(ii). The above inequalities therefore hold as equalities and the theorem is proved. □

Finally we show that the proximal term is useless if r is too small.

Theorem 3.12. *For $r \geqslant 0$ small enough, the global minimizers of $\tilde{\varphi}_r$ are the solutions of the relaxed problem* (3.10) *with $r = 0$.*

Proof. For any $y \in \hat{F}_\infty$, define the number

$$r(y) := \inf\{r > 0 : y \text{ minimizes (globally) } \tilde{\varphi}_r\}.$$

This is a nonnegative number, possibly $+\infty$. Then define

$$\bar{r} := \min\left\{r(y) : y \in \hat{F}_\infty, r(y) > 0\right\};$$

note that $\bar{r} > 0$ because $\hat{F}_\infty$ is a finite set.

Fix $r < \bar{r}$ and let $\bar{y}$ be an arbitrary global minimizer of $\tilde{\varphi}_r$; then $\bar{y} \in \hat{F}_\infty$ and, by definition of $\bar{r}$, $r(\bar{y})$ cannot be positive: $r(\bar{y}) = 0$. This means that there exists a sequence $r_\ell \downarrow 0$ such that $\bar{y}$ is a global minimizer of $\tilde{\varphi}_{r_\ell}$. In view of Proposition 3.7:

$$g^{**}_{r_\ell}(x) - 2r_\ell \bar{y}^\top x \geqslant g^{**}_{r_\ell}(\bar{y}) - 2r_\ell \|\bar{y}\|^2 \tag{3.12}$$

for all x feasible in (3.10)(a) and $\ell = 1, 2, \ldots$

When $r_\ell \downarrow 0$, it is mere technicalities to establish the pointwise convergence of biconjugates (form convex combinations $\sum_k \alpha_{\ell k} x_{\ell k}$; the α's and x's are bounded and it suffices to pass to the limit for each k). Then, passing to the limit in (3.12) shows that $g^{**}(x) \geqslant g^{**}(\bar{y})$. □

3.4. An Example

As an illustrative example, consider the problem with two binary variables x^1, x^2 and one constraint

$$\min_{x \in F} \{x^1 + 4x^2\}, \quad \text{where } F := \left\{x \in Z := \{0,1\}^2 \,:\, x^1 + 2x^2 \geqslant 2\right\}. \tag{3.13}$$

The feasible points are $x^* = (0, 1)$ (the optimal solution, with objective value $g(x^*) = 4 = v^*$) and $e := (1, 1)$, so that

$$\varphi_r(y) = \min\{4 + r\|y - x^*\|^2, 5 + r\|y - e\|^2\}.$$

Working out the calculations, we obtain

$$\varphi_r(y) = \begin{cases} 4 + r\|y - x^*\|^2 & \text{if } (e - x^*)^\top y \leqslant (1 + r)/2r, \\ 5 + r\|y - e\|^2 & \text{otherwise}. \end{cases}$$

First observe that $(e - x^*)^\top x^* = 0$ is always smaller than $(1 + r)/2r$, hence $y = x^*$ is always a local minimum. Now consider three cases:

- If $(1 + r)/2r < 1$ (i.e. $r > 1$) then $y = e$ is another local minimum – Theorem 2.2 (i).
- If $(1 + r)/2r > 1$ (i.e. $r < 1$), this latter local minimum vanishes – Theorem 2.2 (ii).

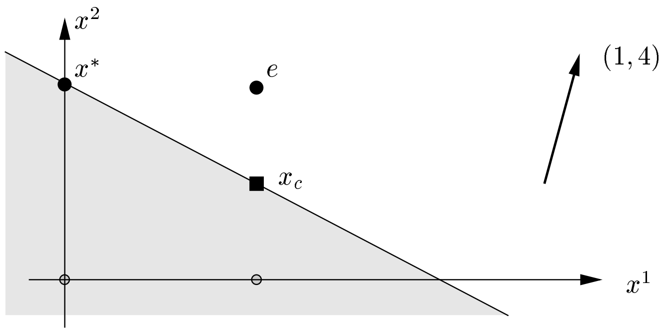

Fig. 3.2. An illustrative example

– We leave it to the reader to check the case $(1+2r)/2r = 1$, and in particular to see why $y = e$ is not a local minimum.

Now we study $\tilde{\varphi}_r$. The relaxed primal solution (i.e. solving (3.10) with $r = 0$) is $x_c = (1, 1/2)$, with objective value $g(x_c) = 3 = v_c$. As a function defined over the whole of $\mathbb{R}^2$, g is linear, hence $\bar{r}_g = 0$ (see Lemma 3.5). Observe that $\hat{F}_r = \{x^*, e, x_c\}$ for all $r \geqslant 0$. We have from (3.10)(a)

$$\tilde{\varphi}_r(y) = r\|y\|^2 + \min\left\{g_r^{**}(x) - 2ry^\top x : x \in \{x^*, e, x_c\}\right\},$$

knowing that

$$g_r^{**}(x^*) = g_r(x^*) = 4 + r - 2ry^2\,,$$
$$g_r^{**}(e) = g_r(e) = 5 + 2r - 2r(y^1 + y^2)$$

(both x^* and e lie in Z), and it is easy to see that

$$g_r^{**}(x_c) = \frac{g_r(1,0) + g_r(e)}{2} = 3 + \frac{3r}{2} - r(2y^1 + y^2)\,.$$

Calculations are left to the reader; the final results are as indicated in Table 3.1. Note:

– $\min \tilde{\varphi}_r = \min\{3 + r/4, 4\}$ which, as predicted by Lemma 1.2(i)), is nondecreasing;
– $\min \tilde{\varphi}_r > 3 = v_c$ for all $r > 0$ (Proposition 3.9);
– each point of the feasible set $\{x^*, e\}$ becomes a local minimum of $\tilde{\varphi}_r$ for $r \geqslant 4$ (Proposition 3.10);
– x^* becomes global minimum of $\tilde{\varphi}_r$ for $r \geqslant 4$ (Theorem 3.11).

r	0	1/3		4			$+\infty$
local mins.	x_c	x_c	x^*	x_c	x^*	e	
$\tilde{\varphi}_r$ values	$3 + r/4$	$3 + r/4$	4	$3 + r/4$	4	5	

Table 3.1. Local minima of $\tilde{\varphi}_r$.

3.5. The Relaxed Proximal Algorithm

Consider now the proximal algorithm (1.10) to minimize $\tilde{\varphi}_r$. It needs a black box to compute $\tilde{\varphi}_r(y)$ for a given $y \in \mathbb{R}^n$. Of course this is done by some optimization process, which produces an $x(y)$ solving one of the "equivalent" problems (3.4) or (3.10).

Then we do the following.

Algorithm 3.13 (Implementable proximal algorithm I) *Choose* $r > 0$.
STEP 0. *Take* y_1 *arbitrary in* $\mathbb{R}^n$*; set* $k = 1$.
STEP 1. *Call the black box to obtain* $x(y_k)$.
STEP 2. *If* $x(y_k) = y_k$ *then stop.*
STEP 3. *Set* $y_{k+1} = x(y_k)$*, replace* k *by* $k+1$ *and go to Step 1.* □

Convergence properties of this algorithm are reminiscent of Theorem 2.4:

Theorem 3.14. *The sequence* y_k *generated by the above algorithm has some cluster point* y^*.

If $r \geqslant \bar{r}_g$*, any such cluster point lying in* Z *is feasible in* (3.1)*, and is a local minimum of* $\varphi_{r'}$ *of* (3.3)*, for any* $r' > r$.

Proof. The first statement is true because y_k varies in $\operatorname{co} Z$, a compact set.

Then take a cluster point $y^* \in Z$. By construction, each $y_k = x(y_{k-1})$ satisfies $Ay_k \geqslant c$; hence $y^* \in F \subset \hat{F}_r$ (Lemma 3.5 (iii)). Then proceed as in Theorem 1.7 (iii) (being convex, the function γ_r is *continuous* on $\operatorname{co} Z$): for all $x \in \hat{F}_r$,

$$\gamma_r(x) + r\|x - y^*\|^2 \geqslant \gamma_r(y^*) = g_r^{**}(y^*) - r\|y^*\|^2 = g(y^*)\,,$$

where the last equality comes from Lemma 3.5 (i).

Take in particular $x \in F$; then $x \in Z$, hence $\gamma_r(x) \leqslant g_r(x) - r\|x\|^2 = g(x)$: we have in fact proved

$$g(x) + r\|x - y^*\|^2 \geqslant g(y^*) \quad \text{for all } x \in F\ .$$

The rest follows from Proposition 1.4. □

Observe that nothing guarantees $y^* \in Z$: actually, y^* may well be in the parasitic set $\hat{F}_r \backslash F$. Since $\hat{F}_r$ is finite, the algorithm would stop at Step 2 if we could guarantee $x(y_k) \in \hat{F}_r$ for each k; but it is not clear how this can be done:

- Either we solve (3.4) by some dual algorithm (i.e. by column generation, see [11, §5.2]); this produces solutions of (3.10)(c), which need *not* lie in $\hat{F}_r$.
- To obtain for sure an extreme point, (3.10) should be solved by a simplex-like algorithm; but this supposes a linear objective $g(x) = b^\top x$ and a close description of the polyhedron $\operatorname{co} Z$ – at least of its intersection with $Ax \geqslant c$. In case (3.2), for example, this essentially means that the so-called integrality property is satisfied, or equivalently that nothing is changed if all constraints are dualized: $Z = \{0,1\}^n$.

As a conclusion, let us compare with the situation in §2.

(i) Replacing φ_r by $\tilde{\varphi}_r$ has a substantial price. The proximal algorithm produces iterates y_k which need no longer be feasible (say $y_k \in \hat{F}_r \backslash F$) and the sequence $g(y_k)$ is no longer decreasing. Instead of producing a feasible point for sure (Theorem 2.4), we may land at some parasitic point in $\hat{F}_r \backslash F$, including a relaxed solution x_c, solving (3.10)(b).

(ii) Taking a small r was recommended in §2; this is now unwise since it will probably produce a relaxed solution (Theorem 3.12). A relaxed solution x_c will even be produced for sure if Algorithm 3.13 is initialized on $y_1 = x_c$.

(iii) The only way to escape from x_c is to increase r. However this is dangerous, since the algorithm might produce a bad feasible point of (3.1) (Proposition 3.10); besides, it is not sure that larger values of r will eventually eliminate x_c from the set of local minima.

(iv) Once again, global minimization of $\tilde{\varphi}_r$ would eliminate any problem; yet we should also take r large enough (Theorem 3.11).

3.6. Case of 0-1 Variables

This section considers the particular case where $Z = \{0,1\}^n$ in (3.1). An interesting property with the present unit cube Z is that all of its points are extreme in $\operatorname{co} Z$. As a result, g can be extended by a convex function to the whole of $\operatorname{co} Z$, the results of the previous sections can therefore be reproduced with $\bar{r}_g = 0$:

Proposition 3.15. *For any $x \in \{0,1\}^n$, $g^{**}(x) = g(x)$.*

Proof. Just apply Remark 3.3, observing that a nontrivial convex combination of 0-1 points cannot be 0-1. □

We now turn to a variant of the proximal algorithm, which can also be used in the present particular case. The trouble with the convexification procedure of §2 is that the quadratic term can by no means convexify the objective function f of (2.2). In fact, the original motivation for [1] was to treat "ordinary" (continuous) nonlinear programming problems, whose Lagrangian's Hessian could be made positive definite, simply by adding a big enough diagonal $r\mathrm{I}$ to it. For a strict application of this idea, we extend g to the whole of $[0,1]^n$ by a convex function h – this is possible thanks to Proposition 3.15, for example with $h := g^{**}$. Then (3.1) takes the form

$$\min h(x)\,, \quad x \in Z := \{0,1\}^n\,, \quad Ax \geqslant c\,, \tag{3.14}$$

the feasible set being

$$F := \{x \in \{0,1\}^n : Ax \geqslant c\}\,. \tag{3.15}$$

We still denote by v^* and v_c the optimal values of (3.14) and of its relaxation, as in (3.5).

Now we penalize the integrality constraints by the factor

$$p(x) := e^\top x - \|x\|^2\,, \quad \text{where } e := (1,\dots,1)\,. \tag{3.16}$$

Taking a large parameter $\pi > 0$, we transform (3.14) to

$$\min\,[h(x) + \pi p(x)]\,, \quad x \in Q \subset \operatorname{co} Z = [0,1]^n\,, \tag{3.17}$$

Q being a closed convex polyhedron containing F. The following result confirms the general belief that (3.17) is just another form of (3.14) if the penalty is big enough. We state this result in a form allowing a large flexibility for Q – of course, the most natural choice is to simply change $\{0,1\}$ to $[0,1]$ in (3.14), which reproduces the feasible set of (3.10)(c).

Theorem 3.16. *Assume in* (3.17) *that the objective function $h + \pi p$ is concave. Then*

(i) $h + \pi' p$ is strictly concave for any $\pi' > \pi$.

Let $h + \pi p$ be strictly concave and assume that the feasible set Q contains the feasible set F of (3.14), (3.15). *Then:*

(ii) The optimal value of (3.17) *is not larger than the optimal value v^* of* (3.14). *Any optimal solution of* (3.17) *lying in F is also optimal for* (3.14).

(iii) Assume further that Q does not contain any other 0-1 point than those in F. For π large enough, any optimal solution of (3.17) *lies in F. As a result:*

- *the optimal value of* (3.17) *is v^*,*
- *the sets of optimal solutions in* (3.14) *and* (3.17) *coincide.*

Proof. (i) Obvious since $h + \pi' p = (h + \pi p) + (\pi' - \pi)p$ and p is strictly concave.

(ii) Just observe that (3.17) is a *relaxation* of (3.14): both their objective functions coincide on the feasible set $F \subset \{0,1\}^n$ of the latter.

(iii) From (ii), v^* is an upper bound for the optimal value of (3.17). With $h + \pi p$ strictly concave on $Q \subset [0,1]^n$, the feasible set in (3.17) can be restricted to the set $\operatorname{ext} Q$ of its extreme points. Denote by $\bar{F} := \operatorname{ext} Q \backslash F$ the set of parasitic points. Set $\delta := \min_{x \in \bar{F}} p(x)$ and note that $\delta > 0$ because $p > 0$ on the finite set $\bar{F}$.

Increase π if necessary so that $\pi > (v^* - \min_Q h)/\delta$, let $x^* \in \operatorname{ext} Q$ solve (3.17) and assume $x^* \in \bar{F}$:

$$v^* \geqslant h(x^*) + \pi p(x^*) \geqslant \min_Q h + \pi\delta > v^*\,,$$

a contradiction. Thus, any x_π optimal in (3.17) lies in F. In view of (ii), x_π solves (3.14); and, since p is 0 on F, the optimal value of (3.17) is v^*. Conversely, any x^* optimal in (3.14) is feasible in (3.17) and has $h(x^*) + \pi p(x^*) = h(x^*) = v^*$; thus x^* is optimal in (3.17). □

Now the proximal convexification procedure can be applied to (3.17) instead of (3.14): we define

$$\psi_{\pi r}(y) := \min_{x \in Q} \left[h(x) + \pi p(x) + r\|y - x\|^2\right] \tag{3.18}$$

and all the results of §1 apply, $\psi_{\pi r}$ playing the role of φ_r. Observe that the minimand $f_r(\cdot, y) = h(\cdot) + \pi\, p(\cdot) + r\|\cdot - y\|^2$ in (3.18) is convex for $r \geqslant \pi$: in contrast with φ_r, the computation of $\psi_{\pi r}$ can be made "easy". By contrast, its global minimization is not straightforward (of course). Assume for example h is linear; then the Hessian of $f_r(\cdot,\cdot)$ is $2\begin{pmatrix}(r-\pi)\mathrm{I} & -r\mathrm{I}\\ -r\mathrm{I} & r\mathrm{I}\end{pmatrix}$ (I is the identity in $\mathbb{R}^n$), which is never positive semidefinite[4].

Algorithm 3.17 (Implementable proximal algorithm II) *A black box is assumed available to solve* (3.18) *for given* y. *Take* $r \geqslant \pi$.
STEP 0. *Take* y_1 *arbitrary in* $\mathbb{R}^n$; *set* $k = 1$.
STEP 1. *Call the black box to obtain* $x(y_k) \in Q$.
STEP 2. *If* $x(y_k) = y_k$ *then stop.*
STEP 3. *Set* $y_{k+1} = x(y_k)$, *replace* k *by* $k+1$ *and go to Step 1.* □

Theorem 3.18. *The sequence* y_k *generated by the above algorithm has some cluster point. Any such cluster point is a local minimum of* $\psi_{\pi r'}$, *for any* $r' > r$.

Proof. Use Theorem 1.7 (iii) (h and p are continuous) and Proposition 2.1. □

The present variant turns out to bring little with respect to §3.1:

Proposition 3.19. *Call* $\hat{Z}$ *the feasible set in* (3.10)*(c) (Z being* $\{0,1\}^n$*).*
(i) If $Q \supset \hat{Z}$, *then* $\psi_{rr}(y) \leqslant \tilde{\varphi}_r(y)$ *for all* $y \in \mathbb{R}^n$ *and* $r \geqslant 0$.
(ii) Equality holds if $Q = \hat{Z}$ *and* h *in* (3.14) *is linear.*
(iii) In the latter case, any local minimum of $\tilde{\varphi}_r$ *is a local minimum of* $\psi_{rr'}$ *for all* $r' > r$.

Proof. In view of Remark 3.3, any x feasible in (3.10)(c) is a convex combination of 0-1 points x_k such that

$$g_r^{**}(x) = \sum_k \alpha_k g(x_k) + r\sum_k \alpha_k \|x_k\|^2 = \sum_k \alpha_k h(x_k) + r\sum_k \alpha_k e^\top x_k\,.$$

Using the convexity of h, we therefore have

$$\gamma_r(x) + r\|x-y\|^2 \geqslant h(x) + r\,p(x) + r\|x-y\|^2\,,$$

with equality if h is linear. In the righthand side above, we recognize the minimand in (3.18) (with $\pi = r$): this latter problem being a relaxation of (3.10)(c), (i) and (ii) are proved.

To prove (iii), the local minima of $\tilde{\varphi}_r$ and of ψ_{rr} coincide in the situation (ii); and in view of Lemma 1.2 (iv), the latter are also local minima of $\psi_{\pi r}$. □

Thus, using the minimal value $r = \pi$ in (3.18) can only increase the duality gap (unless Q can be made smaller). In the "standard" case (h linear, $Q = \hat{Z}$), using $r > \pi$ can only increase the set of parasitic points where Algorithm 3.17 stalls.

[4] In this case, convexity of $\psi_{\pi r}$ becomes unlikely: assuming for example that the constraint $x \in Q$ is inactive in (3.18), a brief calculation shows that $\psi_{\pi r}$ is quadratic with Hessian $\frac{-2r\pi}{r-\pi}\mathrm{I}$.

General Conclusion. We have studied in the framework of combinatorial optimization a general convexification procedure (the primal proximal algorithm), assessed by a useful heuristic for some special problem (unit-commitment, see [5]). This procedure gives birth to various conceptual and implementable heuristic algorithms to generate primal solutions, and our results suggest that little is to be expected in theory from the approach. In fact, the ability itself of the approach in producing feasible suboptimal solutions is controversial. From our conclusion of §3.5, such an ability implies a delicate tuning of the proximity parameter, to avoid

Charybdis: being attracted with a small r toward the relaxed solution x_c, as well as

Scylla: jumping with a large r to an uncontrollable feasible point in (2.1).

Yet, even if these two dangers are avoided, the mere production of a feasible point is never guaranteed.

Acknowledgement. We are indebted to an anonymous referee, whose lucid suggestions helped us to considerably enhance the earlier version [3] of this paper.

References

1. D.P. Bertsekas. Convexification procedures and decomposition methods for nonconvex optimization problems. *Journal of Optimization Theory and Applications*, 29:169–197, 1979.
2. D.P. Bertsekas, G.S. Lauer, N.R. Sandell, and T.A. Posberg. Optimal short-term scheduling of large-scale power systems. *IEEE Transactions on Automatic Control*, AC-28:1–11, 1983.
3. A. Daniilidis and C. Lemaréchal. Proximal convexification procedures in combinatorial optimization. RR 4550, Inria, 2002. `www.inria.fr/rrrt/rr-4550.html`.
4. J.M. Danskin. The theory of max-min with applications. *SIAM Journal on Applied Mathematics*, 14(4):641–655, 1966.
5. L. Dubost, R. Gonzalez, and C. Lemaréchal. A primal-proximal heuristic applied to the french unit-commitment problem. Submitted to Mathematical Programming.
6. J.E. Falk. Lagrange multipliers and nonconvex programs. *SIAM Journal on Control*, 7(4):534–545, 1969.
7. S. Feltenmark and K. C. Kiwiel. Dual applications of proximal bundle methods, including Lagrangian relaxation of nonconvex problems. *SIAM Journal on Optimization*, 10(3):697–721, 2000.
8. A.M. Geoffrion. Lagrangean relaxation for integer programming. *Mathematical Programming Study*, 2:82–114, 1974.
9. J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms.* Springer Verlag, Heidelberg, 1993. Two volumes.
10. J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis.* Springer Verlag, Heidelberg, 2001.
11. C. Lemaréchal. Lagrangian relaxation. In M. Jünger and D. Naddef, editors, *Computational Combinatorial Optimization*, pages 115–160. Springer Verlag, Heidelberg, 2001.
12. C. Lemaréchal, F. Pellegrino, A. Renaud, and C. Sagastizábal. Bundle methods applied to the unit-commitment problem. In J. Doležal and J. Fidler, editors, *System Modelling and Optimization*, pages 395–402. Chapman and Hall, 1996.
13. C. Lemaréchal and A. Renaud. A geometric study of duality gaps, with applications. *Mathematical Programming*, 90(3):399–427, 2001.
14. T.L. Magnanti, J.F. Shapiro, and M.H. Wagner. Generalized linear programming solves the dual. *Management Science*, 22(11):1195–1203, 1976.
15. R.T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control*, 12:268–285, 1974.
16. R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis.* Springer Verlag, Heidelberg, 1998.