How the augmented Lagrangian algorithm can deal with an infeasible convex quadratic optimization problem*

A. Chiche[†‡§] and J. Ch. Gilbert[†]

December 24, 2014

Il y eu d'abord de l'incompréhension lors d'une première lecture de « Proximité et dualité dans un espace hilbertien », puis de l'admiration et enfin du respect pour toujours

— À la mémoire de Jean-Jacques Moreau —

This paper analyses the behavior of the augmented Lagrangian algorithm when it deals with an infeasible convex quadratic optimization problem. It is shown that the algorithm finds a point that, on the one hand, satisfies the constraints shifted by the smallest possible shift that makes them feasible and, on the other hand, minimizes the objective on the corresponding shifted constrained set. The speed of convergence to such a point is globally linear, with a rate that is inversely proportional to the augmentation parameter. This suggests us a rule for determining the augmentation parameter that aims at controlling the speed of convergence of the shifted constraint norm to zero; this rule has the advantage of generating bounded augmentation parameters even when the problem is infeasible.

Keywords: Augmented Lagrangian algorithm, augmentation parameter update, closest feasible problem, convex quadratic optimization, feasible shift, global linear convergence, infeasible problem, proximal point algorithm, quasi-global error bound, shifted constraint.

AMS Subject Classification: 49M27, 49M29, 65K05, 90C05, 90C06, 90C20, 90C25.

1 Introduction

We consider the *convex quadratic optimization problem* that we write as follows

$$\begin{cases} \inf_x \, q(x) \\ l \leqslant Ax \leqslant u. \end{cases} \tag{1.1}$$

In that problem, the objective function

$$q : x \in \mathbb{R}^n \mapsto q(x) = g^\mathsf{T} x + \frac{1}{2} x^\mathsf{T} H x \tag{1.2}$$

*INRIA Research Report RR-8583. [hal-01057577]

†INRIA Paris-Rocquencourt, Pomdapi team – BP 105, F-78153 Le Chesnay Cedex (France) – e-mail: `Jean-Charles.Gilbert@inria.fr`.

‡EDF R&D, Osiris departement – 1, avenue du Général de Gaulle – F-92141 Clamart Cedex (France).

§Artelys France, 12 rue du Quatre Septembre, 75002 Paris (France) – e-mail: `Alice.Chiche@artelys.com`.

is convex quadratic (the vector $g \in \mathbb{R}^n$ and the matrix $H \in \mathbb{R}^{n\times n}$ is positive semidefinite) and the constraints are defined by a matrix $A \in \mathbb{R}^{m\times n}$ and bounds l and $u \in \overline{\mathbb{R}}^m$ that must satisfy $l < u$ (we have used the notation $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$). The sign "$^\mathsf{T}$" denotes transposition, so that $u^\mathsf{T} v = \sum_i u_i v_i$ is the Euclidean scalar product of the vectors u and v. Because of the possible infinite value of the components of l and u, we feel it necessary to give a precise definition of the frequently used interval

$$[l, u] := \{y \in \mathbb{R}^m : l \leqslant y \leqslant u\}. \tag{1.3}$$

Since H may vanish, the problem encompasses linear optimization. On the other hand, linear equality constraints, like $Bx = b$, can be expressed in (1.1) by using two inequalities $Bx \leqslant b$ and $-Bx \leqslant -b$, so that the analysis below also covers problems with linear equality constraints.

The augmented Lagrangian (AL) algorithm studied in this paper is defined by first introducing an auxiliary vector of variables $y \in \mathbb{R}^m$ and by rewriting (1.1) as follows

$$\begin{cases} \inf_{(x,y)}\ q(x) \\ Ax = y \\ l \leqslant y \leqslant u. \end{cases} \tag{1.4}$$

Given an *augmentation parameter* $r \geqslant 0$, the *AL function* $\ell_r : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is then defined at $(x, y, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ by

$$\ell_r(x, y, \lambda) = q(x) + \lambda^\mathsf{T}(Ax - y) + \frac{r}{2}\|Ax - y\|^2, \tag{1.5}$$

where here and below $\|\cdot\|$ denotes the ℓ_2-norm. For $r = 0$, one recovers the usual Lagrangian function, relaxing the equality constraints of (1.4) thanks to the *multiplier* or *dual variable* λ. The *AL algorithm* generates a sequence of dual variables $\{\lambda_k\}_{k\in\mathbb{N}} \subseteq \mathbb{R}^m$ as follows. Knowing $r_k > 0$ and $\lambda_k \in \mathbb{R}^m$, the next dual iterate λ_{k+1} is computed by

$$(x_{k+1}, y_{k+1}) \in \arg\min \left\{\ell_{r_k}(x, y, \lambda_k) : (x, y) \in \mathbb{R}^n \times [l, u]\right\}, \tag{1.6}$$

$$\lambda_{k+1} := \lambda_k + r_k(Ax_{k+1} - y_{k+1}), \tag{1.7}$$

where "arg min" denotes the solution set to the associated minimization problem. Next, r_k is updated by a rule that depends on the implementation and to which we pay much attention in this paper. The quadratic optimization problem in (1.6) is called the *AL subproblem*. The algorithm is presented with more details and is further discussed at the end of this section.

This paper can be viewed as a continuation of the work initiated in [17; 2005], in which the *global* linear convergence of the constraint norm to zero is established, when (1.1) is *feasible* and *bounded*. Feasibility means that there is a point $x \in \mathbb{R}^n$ such that $Ax \in [l, u]$ or, equivalently, that $\mathcal{R}(A) \cap [l, u] \neq \varnothing$ (we denote the *range space* of the matrix A by $\mathcal{R}(A) := \{Ax : x \in \mathbb{R}^n\}$). When feasibility occurs, boundedness means that the optimal value of (1.1) is finite. For a quadratic problem like (1.1), these two conditions (feasibility and boundedness) are equivalent to the existence of a solution [25; 1956]. More specifically, it was shown in [17] that when (1.1) has a solution

$$\begin{array}{c}\forall\, \beta > 0, \quad \exists\, L > 0, \quad \operatorname{dist}(\lambda_0, \mathcal{S}_{\rm D}) \leqslant \beta \quad \text{implies that} \\ \forall\, k \geqslant 1, \quad \|Ax_{k+1} - y_{k+1}\| \leqslant \frac{L}{r_k} \|Ax_k - y_k\|,\end{array} \tag{1.8}$$

where $\mathcal{S}_{\rm D}$ denotes the necessarily nonempty set of optimal multipliers associated with the equality constraints of (1.4) and the operator "dist" denotes the Euclidean distance. Computationally, this result is interesting because it allows the AL algorithm to tune the augmentation parameter r_k on the basis of the examined behavior of the constraint norm ratio $\|Ax_{k+1} - y_{k+1}\|/\|Ax_k - y_k\|$, from the very first iteration. In [27], r_k is increased if this ratio is larger than a desired rate of convergence (this rate is easier to prescribe by the solver user than r_k). Let us stress that it is the fact that the constraint norm inequality in (1.8) is valid from the first iteration, not only asymptotically in an unknown neighborhood of the unknown set $\mathcal{S}_{\rm D}$, that makes this tuning possible. Now, when the problem is *infeasible*, i.e., $\mathcal{R}(A) \cap [l, u] = \varnothing$, the constraint norm cannot, of course, tends to zero and the just described rule for tuning r_k makes the augmentation parameter blow up. In that case, the algorithm could stop if r_k exceeds some threshold like in [3], but one understands that (*i*) it is difficult to specify a universal value for such a threshold, (*ii*) a threshold may be difficult to determine for a particular problem by the user of the code, and (*iii*) probably nothing can be said on the approximate solution obtained when the threshold is exceeded.

This paper gives more properties on the AL algorithm when problem (1.1) is *infeasible*. Since, the AL algorithm is equivalent to the *proximal* (*point*) *algorithm* on the dual function [48; 1973], the present contribution is related to the works describing the behavior of the proximal method on monotone inclusion problems without solution [52, 7, 45, 58, 59; 1976-1987], but it goes a little further, by taking advantage of the special structure of the *quadratic* optimization problem (1.1). In particular, the way the changing penalty parameters r_k intervene in the speed of convergence is highlighted.

One assumption is crucial for making the AL algorithm consistent for infeasible problems. Since $[l, u] \neq \varnothing$, it is always possible to find a *shift* $s \in \mathbb{R}^m$ such that the *shifted constraints* $l \leqslant Ax + s \leqslant u$ are feasible for some $x \in \mathbb{R}^n$; let us call such an s a *feasible shift*. The feasible shifts are clearly the vectors in the set

$$\mathcal{S} := [l, u] + \mathcal{R}(A). \tag{1.9}$$

The fundamental assumption of this study is that for some shift $s \in \mathcal{S}$ (or any such feasible shift, as this will be clarified by the comment after proposition 2.5), the *shifted quadratic optimization problem*

$$\left\{\begin{array}{l}\inf_x \, q(x) \\ l \leqslant Ax + s \leqslant u\end{array}\right. \tag{1.10}$$

has a solution. This assumption is essential in the present context because it is equivalent to saying that each AL subproblem (1.6) has a solution whatever are (or, equivalently, for some) $\lambda \in \mathbb{R}^m$ and $r > 0$ (see proposition 2.5 again), so that the AL algorithm is consistent if and only if that fundamental assumption holds.

Since $\mathcal{S}$ is a nonempty closed convex set (it is a convex polyhedron containing $[l, u]$), there is also one and only one smallest shift $\bar{s} \in \mathcal{S}$, which is the projection of zero on $\mathcal{S}$:

$$\bar{s} := \arg\min_{s \in \mathcal{S}} \|s\|. \tag{1.11}$$

Of course, $\bar{s} = 0$ if and only if problem (1.1) is feasible. Problem (1.10) with $s = \bar{s}$ is called in this paper the *closest feasible problem*. It reads

$$\left\{ \begin{array}{l} \inf_x \ q(x) \\ l \leqslant Ax + \bar{s} \leqslant u. \end{array} \right. \tag{1.12}$$

Computing $\bar{s}$ is not easier than computing a solution to a feasible quadratic problem like (1.1), so that this smallest feasible shift is not computed before running the AL algorithm. We will see, however, that in the AL algorithm the following dual function subgradients

$$s_k := y_k - Ax_k \tag{1.13}$$

converge *globally* linearly to $\bar{s}$, in a way similar to (1.8) but with s_k replaced by $s_k - \bar{s}$ in the second line (theorem 3.4). This result is partly due to the fact that $\bar{s}$ is also the smallest subgradient of the dual function δ associated with problem (1.1) (it will be shown in proposition 2.9, indeed, that the set of all subgradients of δ, denoted $\mathcal{R}(\partial\delta)$, is identical to $\mathcal{S}$) and that the AL algorithm tries to find a multiplier $\overline{\lambda}$ such that $\partial\delta(\overline{\lambda})$ contains that smallest subgradient $\bar{s}$.

The minimum shift $\bar{s}$ is not known when the AL algorithm is running, so that it is less straightforward to use that new global linear convergence for updating the parameter r_k, than it was when $\bar{s} = 0$. We propose instead to use the differences $s'_k := s_{k+1} - s_k$, which also converge globally linearly to zero (a known limit point this time!), provided r_k is sufficiently large. Finally, this analysis results (i) in a new update rule for r_k, which maintains bounded the generated sequence of augmentation parameters even for an infeasible problem, hence avoids introducing useless ill-conditionding (section 4.1) and which computes the smallest feasible shift $\bar{s}$ at a global linear speed and (ii) in a new stopping criterion for the AL algorithm, which can detect that a solution to the closest feasible problem has been obtained to the required precision. The new version of the AL algorithm for solving the convex quadratic problem (1.1) is presented in section 4.2.

Another source of motivation for the present work, to add further to [17; 2005], is a result on the minimization of a *strictly* convex quadratic function q subject to *infeasible* linear *equality* constraints $Bx = b$ (Fortin and Glowinski [24; 1982, remark 5.6, page 42] and Glowinski and Le Tallec [30; 1989, remark 2.13, page 65] claim the result without proof; see also [16; 2006, theorem 4.1] for a related result): the primal sequence generated by the AL algorithm converges globally linearly to the solution to the weakly constrained problem

$$\left\{ \begin{array}{l} \inf_x \ q(x) \\ B^{\mathsf{T}}(Bx - b) = 0. \end{array} \right. \tag{1.14}$$

Therefore, this paper can also be viewed as an extension of the Fortin-Glowinski-LeTallec result to the minimization of a *convex* function (*strict* convexity is no longer required) subject to incompatible *inequality* constraints. Without strict convexity, however, the convergence of the entire primal sequence is no longer ensured, so that the presented linear convergence result is related to the constraint values, instead. Another contribution comes from the impact of the values of the penalty parameters r_k on the speed of convergence: the larger the parameters are, the faster the convergence is; this is an expected property of the AL algorithm.

Notation

We denote by $\mathbb{N} := \{0, 1, 2, \dots\}$ the set of nonnegative integers, by $[n_1 : n_2] := \{n_1, \dots, n_2\} = [n_1, n_2] \cap \mathbb{N}$ the set of integers between $n_1 \in \mathbb{N}$ and $n_2 \in \mathbb{N}$, by $\mathbb{R}$ the set of real numbers, and we set $\mathbb{R}_+ := \{t \in \mathbb{R} : t \geqslant 0\}$, $\mathbb{R}_- := -\mathbb{R}_+$, and $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$.

The notation and concepts of convex analysis that we employ are standard [46, 34, 6]. Let $\mathbb{E}$ be a finite dimensional vector space (below, $\mathbb{E}$ is some $\mathbb{R}^p$). The *asymptotic cone* of a nonempty closed convex set $C \subseteq \mathbb{E}$ is denoted by $C^\infty := \{d \in \mathbb{E} : C + d \subseteq C\}$. We denote by $\mathcal{I}_S$ the *indicator function* of a set $S \subseteq \mathbb{E}$: $\mathcal{I}_S(x) = 0$ if $x \in S$, $\mathcal{I}_S(x) = +\infty$ if $x \notin S$. The *domain* of a function $f : \mathbb{E} \to \overline{\mathbb{R}}$ is defined and denoted by $\operatorname{dom} f := \{x \in \mathbb{E} : f(x) < +\infty\}$ and its *epigraph* by $\operatorname{epi} f := \{(x, \alpha) \in \mathbb{E} \times \mathbb{R} : f(x) \leqslant \alpha\}$. As in [34], $\operatorname{Conv}(\mathbb{E})$ is the set of functions $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ that are convex (i.e., $\operatorname{epi} f$ is convex) and proper (i.e., $\operatorname{epi} f \neq \varnothing$); while $\overline{\operatorname{Conv}}(\mathbb{E})$ is the subset of $\operatorname{Conv}(\mathbb{E})$ of those functions f that are also closed (i.e., $\operatorname{epi} f$ is closed).

Suppose now that $\mathbb{E}$ is endowed with a scalar product denoted by $\langle \cdot, \cdot \rangle$ (below, $\langle \cdot, \cdot \rangle$ is the standard Euclidean scalar product of some $\mathbb{E} = \mathbb{R}^p$). The *normal cone* to a convex set $C \subseteq \mathbb{E}$ at $x \in C$ is denoted by $N_C(x) := \{\nu \in \mathbb{E} : \langle \nu, y - x \rangle \leqslant 0, \text{ for all } y \in C\}$. The Fenchel *conjugate* of $f \in \operatorname{Conv}(\mathbb{E})$ is the function $f^* \in \overline{\operatorname{Conv}}(\mathbb{E})$ defined at $s \in \mathbb{E}$ by $f^*(s) = \sup\{\langle s, x \rangle - f(x) : x \in \mathbb{E}\}$. The *biconjugate* f^{**} of f is the conjugate of f^*; its value at $x \in \mathbb{E}$ is given by $f^{**}(x) = \sup\{\langle s, x \rangle - f^*(s) : s \in \mathbb{E}\}$; it is known that $f^{**} = f$ if and only if $f \in \overline{\operatorname{Conv}}(\mathbb{E})$. The *subdifferential* at $x \in \mathbb{E}$ of $f \in \operatorname{Conv}(\mathbb{E})$ is the set denoted by $\partial f(x) := \{s \in \mathbb{E} : f(x) + f^*(s) = \langle s, x \rangle\}$; it is known that the multifunction $x \mapsto \partial f(x)$ is monotone, i.e., $\langle s_2 - s_1, x_2 - x_1 \rangle \geqslant 0$ whenever for $i = 1, 2$, $x_i \in \mathbb{E}$, and $s_i \in \partial f(x_i)$. The *range space* of ∂f is denoted by $\mathcal{R}(\partial f) := \cup\{\partial f(x) : x \in \mathbb{E}\}$. The *orthogonal projector* on $[l, u]$ is denoted by $P_{[l,u]}$.

The standard augmented Lagrangian algorithm

We conclude this introduction by setting forth precisely the classical AL algorithm that is analyzed in this paper. The algorithm will be rewritten in section 4.2 in a version that incorporates the results of this paper and only differs on the stopping criterion (step 3) and on the way of updating the augmentation parameter (step 4). It is described below as though computation were done in exact arithmetic.

STANDARD AL ALGORITHM to solve (1.1)

Initialization: choose $\lambda_0 \in \mathbb{R}^m$ and $r_0 > 0$.
Repeat for $k = 0, 1, 2, \dots$

1. If the feasible problem

$$\min_{(x,y)\in\mathbb{R}^n\times[l,u]} \ell_{r_k}(x,y,\lambda_k) \tag{1.15}$$

has no solution, exit with a direction $d \in \mathbb{R}^n$ such that

$$g^{\mathsf{T}} d < 0, \qquad Hd = 0, \qquad \text{and} \qquad Ad \in [l,u]^{\infty}. \tag{1.16}$$

Otherwise, denote a solution to (1.15) by (x_{k+1}, y_{k+1}).

2. Update the multiplier

$$\lambda_{k+1} = \lambda_k + r_k(Ax_{k+1} - y_{k+1}). \tag{1.17}$$

3. Stop if

$$Ax_{k+1} = y_{k+1}. \tag{1.18}$$

4. Choose a new augmentation parameter: $r_{k+1} > 0$.

This algorithm deserves some comments.

1. It is shown in proposition 2.5 below that if the AL subproblem (1.15) has no solution, then the closest feasible QP is unbounded and the subproblem (1.15) has no solution, whatever is $\lambda_k \in \mathbb{R}^m$ and $r_k > 0$. Therefore this situation is detected at the very first AL iteration.
2. The fact that a direction $d \in \mathbb{R}^n$ satisfying (1.16) can be found when the AL subproblem has no solution is a consequence of lemma 2.2 below; see remark 2.3 (*iii*). Such a direction is useful when the QP solver is used within the SQP algorithm (see part III in [5] and [29, 35], for instance).
3. The AL subproblem (1.15) may have many solutions (x_{k+1}, λ_{k+1}). Despite that fact, the next multiplier λ_{k+1} is uniquely determined by (1.17). This is discussed after lemma 2.4.
4. Some implementations of the AL algorithm update λ_k with more flexibility than in formula (1.17), for example by taking $\lambda_{k+1} = \lambda_k + \xi_k r_k(Ax_{k+1} - y_{k+1})$, with ξ_k in a compact subset of the open interval $]0, 2[$ (see for example [21; 2012, proposition 11]). The compatibility of this flexibility with our analysis has not been explored.
5. The stopping criterion in step 3 is only valid if the QP (1.1) is feasible, since otherwise $Ax_{k+1} = y_{k+1} \in [l,u]$ cannot be satisfied. The proposed stopping criterion is based on the fact that, when the QP is feasible, a pair (x_{k+1}, y_{k+1}) satisfying (1.18) at this stage of the algorithm is necessarily a solution to (1.4). This stopping criterion will be modified for dealing with infeasible problems in the revised AL algorithm presented in section 4.2.
6. The update of the augmentation parameter in step 4 largely depends on the implementation. The rule proposed in [17] will be adapted to infeasible problems in the revised AL algorithm of section 4.2.

The AL algorithm has a long history that cannot be retraced here. The minimum is certainly to mention that it was introduced for equality constrained nonlinear optimization problems by Hestenes and Powell [33, 42; 1969], and extended to inequality constrained problems by Rockafellar, Buys, Arrow, Gould, and Howe [47, 11, 49, 1, 50; 1971-74]. More recently, its properties when it solves more structured problems have been investigated: linear optimization problems are considered in [41, 31; 1972-1992], quadratic optimization problems in [18, 20, 19, 17, 16, 26; 1999-2008], SDP problems in [37, 38, 32, 61; 2004-2010], and cone constrained optimization problems in [54; 2004].

2 Problem structure

2.1 On quadratic optimization

We quote in this section two results on a quadratic optimization problem, slightly more general than (1.1), namely

$$\inf_{x\in X} q(x), \tag{2.1}$$

where q is the quadratic function (1.2) and X is a convex polyhedron. This generality simplifies the proof of proposition 2.2. Both results are useful in the subsequent analysis. The first one recalls the famous characterization of the existence of a solution established by Frank and Wolfe [25; 1956, appendix (i)], which does not require convexity. The second one requires the convexity of the objective q (see remark 2.3 (*ii*)) and characterizes the unboundedness of (1.1) in terms of the existence of a direction d that has interesting theoretical and numerical properties.

We denote by $\operatorname{val}(P) \in \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ the optimal value of an optimization problem (P), with the convention that $\operatorname{val}(P) = +\infty$ if (P) is an *infeasible minimization* problem. We say that a feasible minimization problem (P) is *unbounded* if $\operatorname{val}(P) = -\infty$ and *bounded* if $\operatorname{val}(P) > -\infty$.

Lemma 2.1 (solvable QP) *Consider problem (2.1) with a quadratic objective q and a polyhedral feasible set X. Then this problem has a solution if and only if $\operatorname{val}((2.1)) \in \mathbb{R}$ (i.e., problem (2.1) is feasible and bounded).*

Lemma 2.2 (unbounded convex QP) *Consider problem (2.1) with a convex quadratic objective q and a nonempty polyhedral feasible set X. Then this problem is unbounded if and only if there is a direction $d \in \mathbb{R}^n$ such that*

$$g^{\mathsf{T}} d < 0, \qquad Hd = 0, \qquad \text{and} \qquad d \in X^{\infty}. \tag{2.2}$$

Proof. [$\Leftarrow$] It is clear that the conditions (2.2) imply the unboundedness of the feasible problem (1.1) since, given an arbitrary point $x_0 \in X \neq \varnothing$, the points $x_k = x_0 + kd$ with $k \in \mathbb{N}$ are in X (definition of X^∞) and $q(x_k) = q(x_0) + k(g^{\mathsf{T}} d) \to -\infty$ when $k \to \infty$.

[$\Rightarrow$] When the problem is unbounded, there is a sequence $\{x_k\}$ of feasible points such that $q(x_k) \to -\infty$. By the continuity of q, the sequence $\{x_k\}$ must be unbounded. Extracting a subsequence if necessary, one can assume that $x_k/\|x_k\|$ converges to some unit norm vector v. This one necessarily satisfies

$$g^{\mathsf{T}} v \leqslant 0, \qquad Hv = 0, \qquad \text{and} \qquad v \in X^{\infty}. \tag{2.3}$$

Indeed, the first condition is obtained by taking the limit in $g^{\mathsf{T}} x_k/\|x_k\| \leqslant q(x_k)/\|x_k\|$ [since $H \succcurlyeq 0$] $\leqslant \gamma/\|x_k\|$ [since $q(x_k) \leqslant \gamma$ for some constant $\gamma \in \mathbb{R}$]; the second condition is obtained by taking the limit in $q(x_k)/\|x_k\|^2 \leqslant \gamma/\|x_k\|^2$, which yields $v^{\mathsf{T}} Hv \leqslant 0$ and subsequently $Hv = 0$ by the positive semidefiniteness of H; and the third condition results from $x_k \in X$, $\|x_k\| \to \infty$, and $x_k/\|x_k\| \to v$, which imply that $v \in X^{\infty}$ [2; definition 2.1.2].

We pursue by induction on the dimension of X (i.e., the dimension of its affine hull $\operatorname{aff} X$), taking inspiration from the proof of lemma 2.1 by Franck and Wolfe [25; 1956, appendix (i)]. Let V be the vector subspace parallel to $\operatorname{aff} X$ and denote by w a vector such that $\operatorname{aff} X = w + V$. It is clear that $v \in V$.

- If $\dim X = 1$, (2.2) is satisfied with $d = v$, since otherwise $g^{\mathsf{T}} v$ would vanish by (2.3) and, for any $x \in X$, $q(x)$ would be the constant $g^{\mathsf{T}} w + \frac{1}{2} w^{\mathsf{T}} Hw$, contradicting the fact that problem (1.1) is unbounded.
- Suppose now that the conditions in (2.2) hold when $\dim X < p$ for some $p \in [2:n]$ and let us prove these conditions when $\dim X = p$. If $g^{\mathsf{T}} v < 0$, (2.3) shows that (2.2) is satisfied with $d = v$. Otherwise $g^{\mathsf{T}} v = 0$ and the function q is constant along the direction v (same argument as in the first point). There are now two complementary subcases to consider.

  If $x'_k := x_k - (v^{\mathsf{T}} x_k)v \in X$ for a subsequence of indices $\mathcal{K} \subseteq \mathbb{N}$, then $x'_k \in X' := X \cap \{v\}^{\perp}$ (since $\|v\| = 1$). Furthermore, $q(x'_k) = q(x_k) \to -\infty$, so that the quadratic problem consisting of minimizing q on the convex polyhedron X' is unbounded. Since $\dim X' < \dim X = p$, the induction assumption implies that there exists a direction d such that $g^{\mathsf{T}} d < 0$, $Hd = 0$, and $d \in (X')^{\infty}$. Now, $X' \subseteq X$ implies that $(X')^{\infty} \subseteq X^{\infty}$, so that (2.2) is proven with that d.

  If $x'_k \notin X$ for k larger than some index k_1, then, for each $k \geqslant k_1$, there is an $\alpha_k \in \mathbb{R}$ such that $x''_k := x_k + \alpha_k v$ is on the boundary of X of (1.1). Since that boundary is formed of a finite number of convex polyhedral sets X_i of dimension $< p$ and since $q(x''_k) = q(x_k) \to -\infty$, one of these polyhedron, say X_j, must contain an unbounded subsequence of $\{x''_k\}$ that again satisfies $q(x''_k) \to -\infty$. The conclusion now follows, like before, from the induction assumption since $\dim X_j < p$ and $X_j^{\infty} \subseteq X^{\infty}$. $\square$

If the convex polyhedron reads $X := \{x : Ax \in [l, u]\}$, like in problem (1.1), there holds $X^{\infty} = \{d : Ad \in [l, u]^{\infty}\}$ and the conditions (2.2) becomes

$$g^{\mathsf{T}} d < 0, \qquad Hd = 0, \qquad \text{and} \qquad Ad \in [l, u]^{\infty}. \tag{2.4}$$

A direction satisfying (2.4) is called in this paper an *unboundedness direction* or a *direction of unboundedness*.

Remarks 2.3 (i) Lemma 2.2 no longer holds if the feasible set is an arbitrary closed convex set. For example $\inf\{x_1 : (x_1, x_2) \in X\} = -\infty$ if $X = \{x \in \mathbb{R}^2 : x_2 \geqslant x_1^2\}$, but $X^\infty = \mathbb{R}_+\{d\}$, where $d = (0, 1)$, while $g^\mathsf{T} d = 0$.

(ii) Lemma 2.2 no longer holds if q is nonconvex. For example $\inf\{-x^2 : x \in \mathbb{R}\} = -\infty$ but $g = 0$ so that there is no direction d such that $g^\mathsf{T} d < 0$.

(iii) If we apply lemma 2.2 to the feasible problem (1.15) with $r_k \equiv r > 0$ and $\lambda_k \equiv \lambda$, we see that it has no solution (or, equivalently, problem (1.15) is unbounded) if and only if there is a direction $(d_x, d_y) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$\Big[g + Hx + A^\mathsf{T}\lambda + rA^\mathsf{T}(Ax - y)\Big]^\mathsf{T} d_x - \Big[\lambda + r(Ax - y)\Big]^\mathsf{T} d_y < 0,$$
$$\begin{pmatrix} H + rA^\mathsf{T}A & -rA^\mathsf{T} \\ -rA & rI \end{pmatrix} \begin{pmatrix} d_x \\ d_y \end{pmatrix} = 0, \qquad \text{and} \qquad d_y \in [l, u]^\infty.$$

These conditions are equivalent to (2.4) and $(d_x, d_y) = (d, Ad)$, so that the directions of unboundedness of problem (1.1) can be detected on problem (1.15) in step 1 of the AL algorithm. □

2.2 The dual function

We introduce a *Lagrangian* of problem (1.4) by dualizing its equality constraints. It is the function $\ell : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \mapsto \mathbb{R}$ defined at (x, y, λ) by

$$\ell(x, y, \lambda) = q(x) + \lambda^\mathsf{T}(Ax - y). \tag{2.5}$$

The *dual function* $\delta : \mathbb{R}^m \to \overline{\mathbb{R}}$ associated with problem (1.4) is then defined at λ by

$$\delta(\lambda) := -\inf_{(x,y) \in \mathbb{R}^n \times [l,u]} \ell(x, y, \lambda). \tag{2.6}$$

With the minus sign in front of the infimum, this function is convex, closed, and does not take the value $-\infty$. Therefore,

$$\delta \in \overline{\mathrm{Conv}}(\mathbb{R}^m) \qquad \Longleftrightarrow \qquad \operatorname{dom} \delta \neq \varnothing. \tag{2.7}$$

For $r > 0$, the *Moreau-Yosida regularization* of the dual function δ [39, 34; 1965] is the function $\delta_r : \mathbb{R}^m \to \overline{\mathbb{R}}$ defined at $\lambda \in \mathbb{R}^m$ by

$$\delta_r(\lambda) = \inf_{\mu \in \mathbb{R}^m} \left(\delta(\mu) + \frac{1}{2r}\|\mu - \lambda\|^2\right). \tag{2.8}$$

A fundamental tool to study the properties of the AL algorithm is the following beautiful result by Rockafellar [48; 1973], which is particularized below to the present context; this result is indeed also valid for general convex optimization problems.

Lemma 2.4 (AL and proximality) *Suppose that the dual function δ defined by (2.6) is in $\overline{\mathrm{Conv}}(\mathbb{R}^m)$ and let $r > 0$. Then $\delta_r(\lambda)$ defined by (2.8) verifies*

$$\delta_r(\lambda) = - \inf_{(x,y)\in\mathbb{R}^n\times[l,u]} \ell_r(x,y,\lambda), \tag{2.9}$$

where ℓ_r is the augmented Lagrangian defined in (1.5). Furthermore, the unique solution λ_+ to the problem in the right hand side of (2.8) is linked to an arbitrary solution (x_+, y_+) to the problem in the right hand side of (2.9) by

$$\lambda_+ = \lambda + r(Ax_+ - y_+) \qquad \textit{and} \qquad y_+ - Ax_+ \in \partial\delta(\lambda_+).$$

The unique solution λ_+ to the problem in the right hand side of (2.8) is called the *proximal point* of λ associated with δ and $r > 0$ and is denoted in this paper by

$$\mathrm{prox}_{\delta,r}(\lambda) := \arg\min_{\mu\in\mathbb{R}^m} \left(\delta(\mu) + \frac{1}{2r}\,\|\mu-\lambda\|^2\right). \tag{2.10}$$

Hence, according to lemma 2.4, the multipliers λ_k generated by the AL algorithm satisfy

$$\lambda_{k+1} = \mathrm{prox}_{\delta,r_k}(\lambda_k) \qquad \text{and} \qquad s_{k+1} \in \partial\delta(\lambda_{k+1}), \tag{2.11}$$

where s_k is defined by (1.13). As a result, the multiplier λ_{k+1} computed by the AL algorithm is uniquely determined, although the AL subproblem in (1.6) may have several solutions (x_{k+1}, y_{k+1}). These facts alone show the importance of the dual function in the analysis of the AL algorithm.

To be comprehensive and clear up any ambiguity, we feel it necessary to restate and prove proposition 3.3 from [17] in the present context, in which problem (1.1) may have no solution (infeasibility or unboundedness); in places, we use a different argument (i.e., lemma 2.2), which makes the proof shorter. The proposition establishes a link between properties of three different objects: the nonemptiness of the dual function domain, the solvability of the feasible shifted quadratic problems, and the solvability of the AL subproblems.

Proposition 2.5 (three expressions of the AL subproblem solvability) *Let be given $s \in \mathcal{S} := [l,u] + \mathcal{R}(A)$, $\lambda \in \mathbb{R}^m$, and $r > 0$. Then the following three properties are equivalent:*

(*i*) $\mathrm{dom}\,\delta \neq \varnothing$,

(*ii*) *the feasible shifted quadratic problem (1.10) has a solution,*

(*iii*) *the augmented Lagrangian subproblem in (2.9) has a solution.*

Proof. $[(i) \Rightarrow (iii)]$ Since $\mathrm{dom}\,\delta \neq \varnothing$, $\delta \in \overline{\mathrm{Conv}}(\mathbb{R}^m)$ by (2.7), so that the optimal value $\delta_r(\lambda)$ of the problem in the right hand side of (2.8) is finite. By lemma 2.4, the optimal value of problem in (2.9) is also finite. As a feasible bounded convex quadratic problem, the problem in (2.9) must have a solution (lemma 2.1).

[$(iii) \Rightarrow (ii)$] We proceed by contradiction. Suppose that the feasible problem (1.10) has no solution. Then this problem is unbounded (lemma 2.1) and there is a direction $d \in \mathbb{R}^n$ such that (2.4) holds (lemma 2.2 and $[l-s, u-s]^\infty = [l,u]^\infty$). Now, by applying lemma 2.2 to problem (1.15), we see that the existence of such a direction d implies that problem (1.15) has no solution (remark 2.3 (iii)).

[$(ii) \Rightarrow (i)$] Let $(\bar{x}, (\bar{\lambda}^l, \bar{\lambda}^u))$ be a primal-dual solution to the feasible shifted problem (1.10), where $\bar{\lambda}^l$ [resp. $\bar{\lambda}^u$] is the multiplier associated with the lower [resp. upper] bound. Then $((\bar{x}, A\bar{x}+s), (\bar{\lambda}^l, \bar{\lambda}^u))$ is a primal-dual solution to the optimization problem in (2.6) with $\lambda = \bar{\lambda}^u - \bar{\lambda}^l$. Hence $\delta \not\equiv +\infty$. □

Since $s \in \mathcal{S}$, $\lambda \in \mathbb{R}^m$, and $r > 0$ are common to all the conditions (i)–(iii) of proposition 2.5 and since condition (i) does not depend on that data, once a shifted quadratic problem (1.10) has a solution for some $s \in \mathcal{S}$, it has a solution for any $s \in \mathcal{S}$. For the same reason, once an augmented Lagrangian subproblem in (2.9) has a solution for some $\lambda \in \mathbb{R}^m$ and $r > 0$, it has a solution whatever are $\lambda \in \mathbb{R}^m$ and $r > 0$. Note also that the result no longer holds when $r = 0$: for exemple (i) may not imply (iii) when $r = 0$ ($\operatorname{dom} \delta \neq \varnothing$ does not necessarily imply that $\operatorname{dom} \delta = \mathbb{R}^m$).

We can now precise the general assumption made throughout this paper in the form of three equivalent properties. This equivalence is a consequence of proposition 2.5.

Assumption 2.6 The following equivalent properties hold:

$$\operatorname{dom} \delta \neq \varnothing \quad [\text{this is equivalent to } \delta \in \overline{\operatorname{Conv}}(\mathbb{R}^m)], \tag{2.12}$$

$$\exists\, s \in \mathbb{R}^m : \quad (1.10) \text{ has a solution,} \tag{2.13}$$

$$\forall\, s \in \mathcal{S} : \quad (1.10) \text{ has a solution.} \tag{2.14}$$

In this paper, we are interested in infeasible problems of the form (1.1). The following proposition gives an expression of feasibility in terms of the dual function (2.6), which is instructive to understand how the AL behaves in case it tries to solve an infeasible problem (see the comment after the proof).

Proposition 2.7 (feasibility and dual function) *Suppose that assumption 2.6 holds. Then, the following two properties are equivalent:*

(i) *problem (1.1) is feasible,*

(ii) *the dual function δ is bounded below,*

(iii) $\delta^*(0) < +\infty$.

Proof. [$(i) \Rightarrow (ii)$] When problem (1.1) is feasible, there is some x_0 such that $y_0 := Ax_0 \in [l,u]$. It follows from the definition (2.6) of δ that, for any $\lambda \in \mathbb{R}^m$, $\delta(\lambda) \geqslant -\ell(x_0, y_0, \lambda) = -q(x_0)$; hence δ is bounded below by $-q(x_0) \in \mathbb{R}$.

$[(ii) \Rightarrow (i)]$ Since $\operatorname{dom}\delta \neq \varnothing$ by (2.12), $\delta(\lambda) \in \mathbb{R}$ for some $\lambda \in \mathbb{R}^m$. On the other hand, since $\bar{s}$ defined by (1.11) is the projection of 0 on $\mathcal{S} := [l, u] + \mathcal{R}(A)$, there holds

$$\forall (x, y) \in \mathbb{R}^n \times [l, u] : \quad (y - Ax)^\mathsf{T} \bar{s} \geqslant \|\bar{s}\|^2.$$

Then, for all $t \geqslant 0$:

$$\delta(\lambda - t\bar{s}) = -\inf_{(x,y)\in\mathbb{R}^n\times[l,u]} \Big[q(x) + (\lambda - t\bar{s})^\mathsf{T}(Ax - y)\Big] \leqslant \delta(\lambda) - t\|\bar{s}\|^2.$$

Since δ is bounded below, there must hold $\bar{s} = 0$, i.e., problem (1.1) is feasible.

$[(ii) \Leftrightarrow (iii)]$ The equivalence comes from the fact that $\delta^*(0) = -\inf_{\lambda\in\mathbb{R}^m} \delta(\lambda)$. □

From proposition 2.7, from the proximal interpretation of the AL algorithm given by lemma 2.4, and from the properties of the proximal algorithm, one readily deduces that, for an infeasible problem (1.1) and for a sequence of augmentation parameter r_k satifying $\sum_{k\geqslant 0} r_k = \infty$, the sequence $\{\lambda_k\}$ generated by the AL algorithm is unbounded and $\delta(\lambda_k) \to -\infty$. This observation gives a first picture on the behavior of the AL algorithm when it is used to solve an infeasible problem. More can be said.

The next two propositions aim at highlighting the link between the set $\mathcal{S}$ of feasible shifts and the range of the dual function subdifferential, denoted $\mathcal{R}(\partial\delta)$. These results are "almost valid" for general convex problems, using similar arguments, but with nuances whose description goes beyond the scope of this paper. To avoid making the presentation too cumbersome, we have preferred staying in the domain of convex *quadratic* optimization, although several arguments are also valid for more general convex problems. In the case of convex quadratic problems,

$$\mathcal{S} = \mathcal{R}(\partial\delta), \tag{2.15}$$

provided assumption 2.6 holds. This identity is surprising, since $\mathcal{S}$ only depends on the objects defining the constraint set (here A, l, and u), while δ also depends on the quadratic objective q. The validity of this identity for a general convex problem is briefly discussed after proposition 2.9 below.

We prove (2.15) by means of the *value function* $v : \mathbb{R}^m \to \overline{\mathbb{R}}$ of problem (1.1), which is defined at $s \in \mathbb{R}^m$ by

$$v(s) := \inf\{q(x) : Ax + s \in [l, u],\ x \in \mathbb{R}^n\}. \tag{2.16}$$

The prominent role we give to v in getting (2.15) comes from the fact that, on the one hand, it has a link with $\mathcal{S}$ through the identity

$$\operatorname{dom} v = \mathcal{S},$$

which is easily verified by using the expression (1.9) of $\mathcal{S}$. On the other hand, the value function has also a link with the dual function. Indeed, in convex optimization, it is known and easy to see that the dual function (2.6) can be introduced from the value function by

$$\delta = v^*, \tag{2.17}$$

where v^* denotes the conjugate function of v for the Euclidean scalar product; see [51; theorem 7] or just use the definitions of v, δ, and the conjugate. For a convex quadratic optimization problem satisfying assumption 2.6, the link between v with δ can be reinforced. Taking the conjugate of both sides of (2.17), one gets $\delta^* = v^{**}$. We show in proposition 2.8 below that $\delta^* = v$ or, equivalently, that $v \in \overline{\text{Conv}}(\mathbb{R}^m)$. As highlighted by the proof, this identity rests on the fact that $v(s)$ is obviously the optimal value of the shifted quadratic optimization problem (1.10), that $\delta^*(s)$ is the optimal value of the Lagrangian dual of the same problem, and that there is no duality gap. In other words, the identity $\delta^* = v$ is a compact way of expressing that, whatever is $s \in \mathbb{R}^m$, problem (1.10) and its dual present no duality gap (provided assumption 2.6 holds).

The next result is certainly not original but, by lack of reference, we give it a proof since it will be helpful in proving proposition 2.9 below. The result is related to theorem 11.42 and example 11.43 in [53], although there finiteness of one of the primal or dual optimal values is assumed, which is not the case here (both optimal values may be $+\infty$). The proof makes use of an auxiliary vector of variables $y \in \mathbb{R}^m$ in the shifted quadratic optimization problem (1.10), which then reads

$$\left\{\begin{array}{l} \inf_{x,y}\, q(x) \\ Ax + s = y \\ l \leqslant y \leqslant u. \end{array}\right. \tag{2.18}$$

Proposition 2.8 (no duality gap) *If assumption 2.6 holds, then* $\delta^* = v \in \overline{\text{Conv}}(\mathbb{R}^m)$.

Proof. By assumption (2.12), $\delta \in \overline{\text{Conv}}(\mathbb{R}^m)$, so that $\delta^* \in \overline{\text{Conv}}(\mathbb{R}^m)$ [46; theorem 12.2]. It remains to prove that $\delta^* = v$. We consider two mutually exclusive cases, $v(s) \in \mathbb{R}$ and $v(s) = +\infty$, one of which must occur (since by assumption 2.6, $v(s) > -\infty$).

- If $v(s) \in \mathbb{R}$, then problem (2.18) has a solution $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times [l, u]$ (lemma 2.1). That problem also has an optimal multiplier $\bar{\lambda}$ associated with the affine constraint $Ax+s = y$. Then, the pair $((\bar{x}, \bar{y}), \bar{\lambda})$ is a saddle-point of the Lagrangian $((x, y), \lambda) \mapsto q(x)+\lambda^{\mathsf{T}}(Ax+s-y)$ on $(\mathbb{R}^n \times [l, u]) \times \mathbb{R}^m$, which implies that there is no duality gap:

$$\begin{aligned} &\inf_{(x,y)\in\mathbb{R}^n\times[l,u]}\ \sup_{\lambda\in\mathbb{R}^m} \Big(q(x) + \lambda^{\mathsf{T}}(Ax + s - y)\Big) \\ &\qquad = \sup_{\lambda\in\mathbb{R}^m}\ \inf_{(x,y)\in\mathbb{R}^n\times[l,u]} \Big(q(x) + \lambda^{\mathsf{T}}(Ax + s - y)\Big). \end{aligned}$$

  The left hand side is clearly $v(s)$. The right hand side also reads

$$\sup_{\lambda\in\mathbb{R}^m} \left[s^{\mathsf{T}}\lambda + \inf_{(x,y)\in\mathbb{R}^n\times[l,u]} \Big(q(x) + \lambda^{\mathsf{T}}(Ax - y)\Big)\right] = \sup_{\lambda\in\mathbb{R}^m} \Big[s^{\mathsf{T}}\lambda - \delta(\lambda)\Big] = \delta^*(s).$$

  We have shown that $\delta^*(s) = v(s)$ when $v(s)$ is finite.

- If $v(s) = +\infty$, problem (2.18) is infeasible. Since the dual function of that problem is $\lambda \mapsto \delta(\lambda) - s^{\mathsf{T}}\lambda$, the contrapositive of the implication $(iii) \Rightarrow (i)$ of proposition 2.7 shows that its conjugate at zero, namely $\delta^*(s)$, has the value $+\infty$. $\square$

Note that when assumption 2.6 does not hold, then $\delta \equiv +\infty$ and $\delta^* \equiv -\infty$, while $v(s) = +\infty$ when $s \notin \mathcal{S}$; therefore $\delta^* \neq v$ on the complementary set of $\mathcal{S}$, which may be nonempty (both δ^* and v take the value $-\infty$ on $\mathcal{S}$).

We now show the identity $\mathcal{S} = \mathcal{R}(\partial\delta)$ in (2.15), together with some equivalences, when assumption 2.6 holds. These equivalences, giving various expressions of the fact that λ is a dual solution to problem (2.18), are standard and will be useful below.

Proposition 2.9 (dual subdifferential and feasible shifts) *Suppose that assumption 2.6 holds. Let s and $\lambda \in \mathbb{R}^m$. Then, the following properties are equivalent*

(i) $s \in \partial\delta(\lambda)$,

(ii) $\lambda \in \partial v(s)$,

(iii) $s \in \mathcal{S}$ *and any solution to problem (2.18) minimizes the Lagrangian* $\ell(\cdot,\cdot,\lambda)$ *on* $\mathbb{R}^n \times [l,u]$,

(iv) there is a feasible pair for problem (2.18) that minimizes the Lagrangian $\ell(\cdot,\cdot,\lambda)$ *on* $\mathbb{R}^n \times [l,u]$.

In addition, $\mathcal{S} = \mathcal{R}(\partial\delta)$ *holds.*

Proof. Before proving the equivalences, let us recall that $\delta \in \operatorname{Conv}(\mathbb{R}^m)$ by (2.12), so that

$$s \in \partial\delta(\lambda) \quad \Longleftrightarrow \quad \delta(\lambda) + \delta^*(s) = s^{\mathsf{T}}\lambda. \tag{2.19}$$

$[(i) \Leftrightarrow (ii)]$ Since $\delta \in \overline{\operatorname{Conv}}(\mathbb{R}^m)$ by (2.12), $s \in \partial\delta(\lambda)$ if and only if $\lambda \in \partial\delta^*(s)$ [46; theorem 23.5]. By proposition 2.8, the property $s \in \partial\delta(\lambda)$ is equivalent to $\lambda \in \partial v(s)$.

$[(i), (ii) \Rightarrow (iii)]$ Let $s \in \partial\delta(\lambda)$. By (ii), $s \in \operatorname{dom} v = \mathcal{S}$. Now, let (x_s, y_s) be an arbitrary solution to (2.18). Then

$$\begin{aligned}
\ell(x_s, y_s, \lambda) &= q(x_s) - s^{\mathsf{T}}\lambda \qquad [Ax_s + s = y_s] \\
&= v(s) - s^{\mathsf{T}}\lambda \qquad [\text{definition of } v] \\
&= \delta^*(s) - s^{\mathsf{T}}\lambda \qquad [\text{proposition 2.8}] \\
&= -\delta(\lambda) \qquad [(2.19) \text{ and } s \in \partial\delta(\lambda)] \\
&= \inf_{(x,y)\in\mathbb{R}^n\times[l,u]} \ell(x,y,\lambda) \qquad [\text{definition of } \delta \text{ in (2.6)}].
\end{aligned}$$

This shows the minimality property of (x_s, y_s).

$[(iii) \Rightarrow (iv)]$ This is a clear consequence of the fact that problem (2.18) has a solution when $s \in \mathcal{S}$ and assumption 2.6 holds.

$[(iv) \Rightarrow (i)]$ Let (x_s, y_s) be a feasible point of problem (2.18) with the minimality property mentioned in (iv). Then for any $\mu \in \mathbb{R}^m$:

$$\begin{aligned}
& s^\mathsf{T}\mu - \delta(\mu) \\
&\quad \leqslant \ q(x_s) + \mu^\mathsf{T}(Ax_s - y_s + s) \qquad \text{[definition of } \delta \text{ in (2.6)]} \\
&\quad = \ q(x_s) + \lambda^\mathsf{T}(Ax_s - y_s + s) \qquad \text{[feasibility of } (x_s, y_s)\text{, implying } Ax_s + s = y_s\text{]} \\
&\quad = \ s^\mathsf{T}\lambda + \inf_{(x,y)\in\mathbb{R}^n\times[l,u]} q(x) + \lambda^\mathsf{T}(Ax - y) \qquad \text{[minimality property of } (x_s, y_s)\text{]} \\
&\quad = \ s^\mathsf{T}\lambda - \delta(\lambda) \qquad \text{[definition of } \delta \text{ in (2.6)].}
\end{aligned}$$

Therefore λ minimizes $\mu \in \mathbb{R}^n \mapsto \delta(\mu) - s^\mathsf{T}\mu$, which implies that $s \in \partial\delta(\lambda)$.

$[\mathcal{S} = \mathcal{R}(\partial\delta)]$ The inclusion $\mathcal{R}(\partial\delta) \subseteq \mathcal{S}$ was shown during the proof of "(i), $(ii) \Rightarrow (iii)$". To prove $\mathcal{S} \subseteq \mathcal{R}(\partial\delta)$, let $s \in \mathcal{S}$. By assumption 2.6, problem (2.18) has a primal-dual solution $((x_s, y_s), \lambda_s)$. Hence (x_s, y_s) minimizes $(x, y) \mapsto \ell(x, y, \lambda_s) + s^\mathsf{T}\lambda_s$ on $\mathbb{R}^n \times [l, u]$ and, therefore, also minimizes $(x, y) \mapsto \ell(x, y, \lambda_s)$ on $\mathbb{R}^n \times [l, u]$. By the implication $(iv) \Rightarrow (i)$, $s \in \partial\delta(\lambda_s)$; hence $s \in \mathcal{R}(\partial\delta)$. □

A proof of the identity $\mathcal{S} = \mathcal{R}(\partial\delta)$ can almost be obtained by using general arguments. Note first that for any function $\delta \in \overline{\mathrm{Conv}}(\mathbb{R}^m)$, not necessarily a dual function, there holds

$$\mathrm{ri}(\mathrm{dom}\,\delta^*) \subseteq \mathcal{R}(\partial\delta) \subseteq \mathrm{dom}\,\delta^*,$$

where "ri" denotes the relative interior [46; p. 227]. Taking the closure, one gets $\mathrm{cl}\,\mathcal{R}(\partial\delta) = \mathrm{cl}\,\mathrm{dom}\,\delta^*$. Now, for the dual function δ of problem (1.1), we have by proposition 2.8, $\mathrm{cl}\,\mathcal{R}(\partial\delta) = \mathrm{cl}\,\mathrm{dom}\,v = \mathrm{cl}\,\mathcal{S}$, which would yield $\mathcal{S} = \mathcal{R}(\partial\delta)$ if we knew that $\mathcal{R}(\partial\delta)$ is close (in our case, $\mathcal{S}$ is clearly closed as the sum of two convex polyhedra).

Examples 2.10 ($\boldsymbol{\mathcal{S}}$ not closed) For non-polyhedral constraints, $\mathcal{S}$ may not be closed. Here are two exemples.

1. Consider the nonempty constraint set $\{x \in \mathbb{R} : e^x \leqslant 1\}$. Then, the set of feasible shifts reads $\mathcal{S} := \{s \in \mathbb{R} : \text{there exists an } x \text{ such that } e^x + s \leqslant 1\} = \{s \in \mathbb{R} : s < 1\}$, which is open.
2. Consider the set defined by the linear matrix inequality system of the form $A(X) = b$ and $X \in \mathbb{S}^n_+$, where A maps linearly the space of symmetric matrices $\mathbb{S}^n$ into $\mathbb{R}^m$ and $\mathbb{S}^n_+$ denotes the cone of positive semi-definite matrices. Then, the set of feasible shifts reads $\mathcal{S} = \{s \in \mathbb{R}^m : \text{there is an } X \in \mathbb{S}^n_+ \text{ such that } A(X) + s = b\} = b - A(\mathbb{S}^n_+)$, which may not be closed, since a linear map may transform $\mathbb{S}^n_+$ in a nonclosed set. □

Examples 2.11 (strict inclusion $\boldsymbol{\mathcal{R}(\partial\delta)} \subset \boldsymbol{\mathcal{S}}$) 1. For the convex quadratic problem (1.1) without assumption 2.6, the dual function $\delta \equiv +\infty$. Therefore $\mathcal{R}(\partial\delta) = \varnothing$, while $\mathcal{S} \neq \varnothing$.

2. For a non-quadratic problem, one can have the situation in which $\mathcal{S}$ is closed but strictly larger than $\mathcal{R}(\partial\delta)$. Consider the problem in the single variable $x \in \mathbb{R}$: $\inf\{x : x^2 \leqslant 0\}$. Then $\mathcal{S} := \{s \in \mathbb{R} : x^2 + s \leqslant 0 \text{ for some } x \in \mathbb{R}\} = \{s \in \mathbb{R} : s \leqslant 0\}$. The dual function $\lambda \in \mathbb{R} \mapsto \delta(\lambda) := -\inf\{x + \lambda(x^2 - y) : (x, y) \in \mathbb{R} \times \mathbb{R}_-\}$ verifies

$$\delta(\lambda) = \begin{cases} 1/(4\lambda) & \text{if } \lambda > 0 \\ +\infty & \text{otherwise,} \end{cases} \qquad \text{and} \qquad \partial\delta(\lambda) = \begin{cases} \{-1/(4\lambda^2)\} & \text{if } \lambda > 0 \\ \varnothing & \text{otherwise.} \end{cases}$$

Hence $\mathcal{R}(\partial\delta) = \{s \in \mathbb{R} : s < 0\}$, which is strictly smaller than $\mathcal{S}$. This example and the conditions (*iii*) and (*iv*) of proposition 2.9 highlight the benefit of a dual solution to problem (2.18), which does not exist here when $s = 0$ (the zero feasible shift is precisely the one that is not in $\mathcal{R}(\partial\delta)$). □

The global linear convergence of the AL algorithm is based on the following quasi-global error bound on the dual solution set

$$\mathcal{S}_{\rm D} := \{\lambda \in \mathbb{R}^m : 0 \in \partial\delta(\lambda)\} \tag{2.20}$$

of the feasible QP (1.4) [17; proposition 4.4].

Lemma 2.12 (quasi-global error bound) *Consider problem* (1.4) *with* $H \succcurlyeq 0$ *and suppose that it has a solution. Then*

$$\begin{array}{c} \textit{for any bounded set } \mathcal{B} \subseteq \mathbb{R}^m, \textit{ there is an } L > 0, \textit{ such that} \\ \forall\, \lambda \in \mathcal{S}_{\rm D} + \mathcal{B}, \quad \forall\, s \in \partial\delta(\lambda) : \quad \operatorname{dist}(\lambda, \mathcal{S}_{\rm D}) \leqslant L\,\|s\|. \end{array} \tag{2.21}$$

We use the word *quasi-global* to qualify this error bound since the constant L in (2.21) depends on the bounded set $\mathcal{B}$ and may be infinite (i.e., may not exist) if $\mathcal{B} = \mathbb{R}^m$. This is the case for instance for the feasible problem $\inf_{x\in\mathbb{R}}\{0 : -1 \leqslant 0x \leqslant 1\}$ [17; example 4.3], for which the dual function is $\lambda \in \mathbb{R} \mapsto \delta(\lambda) = |\lambda|$, so that $\mathcal{S}_{\rm D} = \{0\}$ and the last inequality in (2.21) reads $|\lambda| \leqslant L$, which, obviously, cannot hold for all $\lambda \in \mathbb{R}$. The necessity to use a bounded set $\mathcal{B}$ will imply no restriction on the global linear convergence of theorem 3.4, since it will be possible to choose $\mathcal{B}$ such the $\lambda_0 \in \mathcal{S}_{\rm D} + \mathcal{B}$ implies that the next dual iterates $\lambda_k \in \mathcal{S}_{\rm D} + \mathcal{B}$ (proof of lemma 3.3). Now, when problem (1.1) is infeasible, $\mathcal{S}_{\rm D} = \varnothing$, but lemma 2.21 will be used with the dual solution set $\tilde{\mathcal{S}}_{\rm D}$ of the closest feasible problem, introduced in section 2.4.

2.3 The smallest feasible shift

The smallest feasible shift $\bar{s}$ is defined by (1.11) as the smallest element in $\mathcal{S} = [l, u] + \mathcal{R}(A)$ for the Euclidean norm. Clearly, $\bar{s}$ is perpendicular to $\mathcal{R}(A)$, which reads

$$A^{\mathsf{T}}\bar{s} = 0. \tag{2.22}$$

The next lemma gives conditions equivalent to the fact that a pair $(\overline{x}, \overline{y})$ realizes at best the constraint $Ax = y$, in the ℓ_2-norm sense:

$$\min_{(x,y)\in\mathbb{R}^n\times[l,u]} \|Ax - y\|. \tag{2.23}$$

The interest of the conditions in point (*ii*) is that they do not make use of the vector $\bar{s}$, which is unknown when the AL algorithm is trying to solve (1.1). These conditions (*ii*) are a first step in the design of a stopping criterion of the revised version of the AL algorithm, given in section 4.2. The next step is in proposition 2.18.

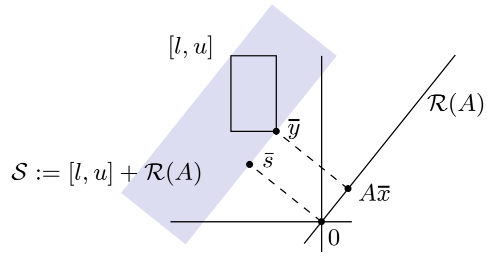

Figure 2.1: Illustration of lemma 2.13

Lemma 2.13 *The following properties of* $(\overline{x}, \overline{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ *are equivalent:*

(i) $\overline{y} - A\overline{x} = \bar{s}$ *and* $\overline{y} \in [l,u]$,

(ii) $A^{\mathsf{T}}(A\overline{x} - \overline{y}) = 0$ *and* $P_{[l,u]}(A\overline{x}) = \overline{y}$,

(iii) $(\overline{x}, \overline{y})$ *is a solution to* (2.23).

Proof. $[(i) \Rightarrow (ii)]$ Since $A^{\mathsf{T}}\bar{s} = 0$ by (2.22), the first identity is a clear consequence of $\overline{y} - A\overline{x} = \bar{s}$ in (i). Now, since $\bar{s}$ is the projection of zero on $[l,u] + \mathcal{R}(A)$, there holds

$$\bar{s}^{\mathsf{T}}(s - \bar{s}) \geqslant 0, \qquad \forall\, s \in [l,u] + \mathcal{R}(A).$$

Choosing $s = y \in [l,u]$, substituting $\bar{s} = \overline{y} - A\overline{x}$, and using the identity $A^{\mathsf{T}}\bar{s} = 0$ yield

$$(\overline{y} - A\overline{x})^{\mathsf{T}}(y - \overline{y}) \geqslant 0, \qquad \forall\, y \in [l,u],$$

which shows the second identity.

$[(ii) \Rightarrow (iii)]$ Using the function $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined at $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$ by $\varphi(x,y) = \frac{1}{2}\|Ax - y\|^2$, the conditions in (ii) can also be written

$$\nabla_x \varphi(\overline{x}, \overline{y}) = 0 \qquad \text{and} \qquad \nabla_y \varphi(\overline{x}, \overline{y})^{\mathsf{T}}(y - \overline{y}) \geqslant 0, \quad \forall\, y \in [l,u].$$

These are the optimality conditions of the convex problem (2.23). Hence $(\overline{x}, \overline{y})$ is a solution to that problem.

$[(iii) \Rightarrow (i)]$ This is because problem (2.23) is equivalent to problem $\inf\{\|s\| : y - Ax = s,\ (x,y) \in \mathbb{R}^n \times [l,u]\} = \inf\{\|s\| : s \in \mathcal{S}\}$, whose solution is $\bar{s}$. Hence (i). $\square$

2.4 The closest feasible problem

Recall that the *closest feasible problem* is the relaxation (1.12) of the possibly infeasible problem (1.1). Using an auxiliary vector $y \in \mathbb{R}^m$, it can be written in one of the two forms

$$\left\{\begin{array}{l} \inf_{(x,y)}\ q(x) \\ Ax + \bar{s} = y \\ l \leqslant y \leqslant u. \end{array}\right. \qquad \text{or} \qquad \left\{\begin{array}{l} \inf_{(x,y)}\ q(x) \\ Ax = y \\ l - \bar{s} \leqslant y \leqslant u - \bar{s}. \end{array}\right. \tag{2.24}$$

That problem is therefore feasible. Of course, if the original problem (1.1) is feasible, $\bar{s} = 0$ and (2.24) is identical to (1.4). Problem (2.24) is the one that the AL algorithm will actually solve, when it is used to solve (1.4). This section gives some properties of that problem.

The dual function $\tilde{\delta} : \mathbb{R}^m \to \overline{\mathbb{R}}$ associated with the closest feasible problem takes at $\lambda \in \mathbb{R}^m$ the value

$$\tilde{\delta}(\lambda) = - \inf_{(x,y)\in\mathbb{R}^n\times[l-\bar{s},u-\bar{s}]} q(x) + \lambda^{\mathsf{T}}(Ax - y) \tag{2.25}$$

$$= - \inf_{(x,y)\in\mathbb{R}^n\times[l,u]} q(x) + \lambda^{\mathsf{T}}(Ax + \bar{s} - y) \tag{2.26}$$

$$= \delta(\lambda) - \bar{s}^{\mathsf{T}}\lambda. \tag{2.27}$$

As a result

$$\partial\tilde{\delta}(\lambda) = \partial\delta(\lambda) - \bar{s}. \tag{2.28}$$

The set of dual solutions to problem (2.24) is denoted by

$$\tilde{\mathcal{S}}_{\mathrm{D}} := \{\bar{\lambda} \in \mathbb{R}^m : 0 \in \partial\tilde{\delta}(\bar{\lambda})\}. \tag{2.29}$$

Here are some other expressions of $\tilde{\mathcal{S}}_{\mathrm{D}}$.

Lemma 2.14 (expressions of $\tilde{\mathcal{S}}_{\mathrm{D}}$) *When assumption 2.6 holds, $\tilde{\mathcal{S}}_{\mathrm{D}}$ is a nonempty closed convex set, which also reads*

$$\tilde{\mathcal{S}}_{\mathrm{D}} = \{\bar{\lambda} \in \mathbb{R}^m : \bar{s} \in \partial\delta(\bar{\lambda})\} = \partial v(\bar{s}). \tag{2.30}$$

Proof. The first equality in (2.30) comes from (2.28) and the second comes from the equivalence $(i) \Leftrightarrow (ii)$ in proposition 2.9. Now, by the identity $\mathcal{S} = \mathcal{R}(\partial\delta)$ in (2.15), $\bar{s} \in \mathcal{S}$ implies the existence of some $\bar{\lambda} \in \mathbb{R}^m$ such that $\bar{s} = \partial\delta(\bar{\lambda})$, so that $\tilde{\mathcal{S}}_{\mathrm{D}} \neq \varnothing$. Since $\tilde{\mathcal{S}}_{\mathrm{D}}$ is the set of minimizers of the closed convex function $\tilde{\delta}$, it is closed and convex. □

The next proposition will be useful to identify some displacement decreasing the distance to $\tilde{\mathcal{S}}_{\mathrm{D}}$. Recall that the *Hadamard product* of two vectors u and $v \in \mathbb{R}^m$ is the vector, denoted $u \cdot v \in \mathbb{R}^m$, having its ith component defined by

$$(u \cdot v)_i = u_i v_i. \tag{2.31}$$

Lemma 2.15 ($\bar{s}$ and $\tilde{\mathcal{S}}_{\mathrm{D}}$) *Suppose that assumption 2.6 holds and let $\bar{\lambda} \in \tilde{\mathcal{S}}_{\mathrm{D}}$. Then*

(i) $\bar{s} \cdot \bar{\lambda} \leqslant 0$,

(ii) if $\bar{s} \cdot (\bar{\lambda} + \alpha\bar{s}) \leqslant 0$ *for some* $\alpha \in \mathbb{R}$, *then* $\bar{\lambda} + \alpha\bar{s} \in \tilde{\mathcal{S}}_{\mathrm{D}}$,

(iii) $-\bar{s} \in \tilde{\mathcal{S}}_{\mathrm{D}}^{\infty}$.

PROOF. [Preliminaries] By $\bar{s} \in \mathcal{S}$ and assumption 2.6, the closest feasible problem (2.24) has a solution, say $(\overline{x}, \overline{y})$. By the assumption $\overline{\lambda} \in \tilde{\mathcal{S}}_{\mathrm{D}}$, the expression (2.30) of $\tilde{\mathcal{S}}_{\mathrm{D}}$ shows that $\bar{s} \in \partial\delta(\overline{\lambda})$. Now, the implication $(i) \Rightarrow (iii)$ of proposition 2.9 indicates that $(\overline{x}, \overline{y})$ minimizes the Lagrangian $\ell(\cdot, \cdot, \overline{\lambda})$ on $\mathbb{R}^n \times [l, u]$.

[(i)] Suppose that $\bar{s}_i > 0$ for some index i (the reasoning is similar when $\bar{s}_i < 0$ and there is nothing to prove when $\bar{s}_i = 0$).

- We first show, by contradiction, that $\overline{y}_i = l_i$. Since $\bar{s} = \arg\min\{\|s\| : s \in [l, u] + \mathcal{R}(A)\}$,

$$(y - Ax - \bar{s})^{\mathsf{T}} \bar{s} \geqslant 0, \qquad \forall\, (x, y) \in \mathbb{R}^n \times [l, u]. \tag{2.32}$$

  If $\overline{y}_i > l_i$, $\overline{y} - \varepsilon e^i = A\overline{x} + \bar{s} - \varepsilon e^i$ is in $[l, u]$ for some $\varepsilon > 0$ (e^i denotes the ith basis vector of $\mathbb{R}^m$). Taking $y = A\overline{x} + \bar{s} - \varepsilon e^i$ and $x = \overline{x}$ in (2.32) yields $-\varepsilon \bar{s}_i \geqslant 0$, a contradiction. Hence $\overline{y}_i = l_i$.
- Now since $(\overline{x}, \overline{y})$ minimizes the Lagrangian $\ell(\cdot, \cdot, \overline{\lambda})$ on $\mathbb{R}^n \times [l, u]$ and since $l < u$, the fact that $\overline{y}_i = l_i$ implies $\overline{\lambda}_i \leqslant 0$. We have shown that $\bar{s}_i \overline{\lambda}_i \leqslant 0$.

[(ii)] We have seen that $(\overline{x}, \overline{y})$ minimizes the Lagrangian $\ell(\cdot, \cdot, \overline{\lambda})$ on $\mathbb{R}^n \times [l, u]$. Suppose that $\alpha \in \mathbb{R}$ is such that $\bar{s} \cdot (\overline{\lambda} + \alpha \bar{s}) \leqslant 0$. The implication $(iv) \Rightarrow (i)$ of proposition 2.9 tells us that to prove that $\overline{\lambda} + \alpha \bar{s} \in \tilde{\mathcal{S}}_{\mathrm{D}}$, which is equivalent to $\bar{s} \in \partial\delta(\overline{\lambda} + \alpha \bar{s})$ by (2.30), we only have to show that $(\overline{x}, \overline{y})$ minimizes

$$(x, y) \mapsto \ell(x, y, \overline{\lambda} + \alpha \bar{s}) = q(x) + (\overline{\lambda} + \alpha \bar{s})^{\mathsf{T}} (Ax - y)$$

on $\mathbb{R}^n \times [l, u]$. By (2.22), the minimization in x is not affected by the new term $\alpha \bar{s}$. As for the minimization in y_i (the minimization in y can be done component by component), we only consider the case when $\bar{s}_i > 0$ (the case $\bar{s}_i < 0$ is similar and, when $\bar{s}_i = 0$, the term in y_i of $\ell(x, y, \overline{\lambda} + \alpha \bar{s})$ is the same as the one of $\ell(x, y, \overline{\lambda})$ so that $\overline{y}_i$ is still a minimizer of $y_i \mapsto \ell(x, y, \overline{\lambda} + \alpha \bar{s})$ on $[l_i, u_i]$). By the proof of (i), we know that $\overline{y}_i = l_i$ in that case, so that it is enough to show that $(\overline{\lambda} + \alpha \bar{s})_i \leqslant 0$, which is indeed verified since $\bar{s} \cdot (\overline{\lambda} + \alpha \bar{s}) \leqslant 0$ by assumption.

[(iii)] Let $\alpha \geqslant 0$. By point (i), $\bar{s} \cdot (\overline{\lambda} - \alpha \bar{s}) \leqslant -\alpha(\bar{s} \cdot \bar{s}) \leqslant 0$. Therefore, by point (ii), $\overline{\lambda} - \alpha \bar{s} \in \tilde{\mathcal{S}}_{\mathrm{D}}$ for all $\alpha \geqslant 0$, meaning that $-\bar{s} \in \tilde{\mathcal{S}}_{\mathrm{D}}^{\infty}$. □

The example below shows that, if $\tilde{\mathcal{S}}_{\mathrm{D}}^{\infty}$ contains the half line $-\mathbb{R}_+ \bar{s}$ (point (iii) of the previous lemma), it is not necessarily reduced to it.

Example 2.16 ($\tilde{\mathcal{S}}_{\mathrm{D}}^{\infty}$ can be an orthant) For the trivial optimization problem with $n = 1$, $m = 2$, $g = 0$, $H = 0$, $A = 0$, $l = (-\infty, -\infty)$, and $u = (-1, -1)$, one finds $\bar{s} = u$ by the definition (1.11) of $\bar{s}$ and $\tilde{\delta} = \mathcal{I}_{\mathbb{R}^2_+}$ by the definition (2.26) of $\tilde{\delta}$, so that $\tilde{\mathcal{S}}_{\mathrm{D}} = \mathbb{R}^2_+ = \tilde{\mathcal{S}}_{\mathrm{D}}^{\infty}$. □

We recall that the prox operator is defined in (2.10).

Lemma 2.17 (distance to $\tilde{\mathcal{S}}_{\mathrm{D}}$) *Suppose that assumption 2.6 holds and let $\lambda \in \mathbb{R}^m$.*

Then the following properties hold:

(*i*) $\operatorname{dist}(\lambda - \alpha\bar{s}, \tilde{\mathcal{S}}_{\mathrm{D}}) \leqslant \operatorname{dist}(\lambda, \tilde{\mathcal{S}}_{\mathrm{D}})$, *when* $\alpha \geqslant 0$,

(*ii*) $\operatorname{prox}_{\delta,r}(\lambda) = \operatorname{prox}_{\tilde{\delta},r}(\lambda - r\bar{s})$,

(*iii*) $\operatorname{dist}(\operatorname{prox}_{\delta,r}(\lambda), \tilde{\mathcal{S}}_{\mathrm{D}}) \leqslant \operatorname{dist}(\lambda, \tilde{\mathcal{S}}_{\mathrm{D}})$.

PROOF. [(*i*)] Let $\tilde{\lambda}$ be the projection of λ on the nonempty closed convex set $\tilde{\mathcal{S}}_{\mathrm{D}}$ (lemma 2.14). For $\alpha \geqslant 0$, $\tilde{\lambda} - \alpha\bar{s} \in \tilde{\mathcal{S}}_{\mathrm{D}}$ (point (*iii*) of lemma 2.15), so that

$$\operatorname{dist}(\lambda - \alpha\bar{s}, \tilde{\mathcal{S}}_{\mathrm{D}}) \leqslant \|(\lambda - \alpha\bar{s}) - (\tilde{\lambda} - \alpha\bar{s})\| = \|\lambda - \tilde{\lambda}\| = \operatorname{dist}(\lambda, \tilde{\mathcal{S}}_{\mathrm{D}}).$$

[(*ii*)] Let $\mu := \operatorname{prox}_{\tilde{\delta},r}(\lambda - r\bar{s})$. Then, $0 \in \partial\tilde{\delta}(\mu) + \frac{1}{r}[\mu - (\lambda - r\bar{s})]$, so that there is some $\tilde{s} \in \partial\tilde{\delta}(\mu)$ such that

$$\mu = \lambda - r(\tilde{s} + \bar{s}).$$

Now $\tilde{s} + \bar{s} \in \partial\tilde{\delta}(\mu) + \bar{s} = \partial\delta(\mu)$ by (2.28), so that $\mu := \operatorname{prox}_{\delta,r}(\lambda)$.

[(*iii*)] By (*ii*),

$$\operatorname{dist}\left(\operatorname{prox}_{\delta,r}(\lambda), \tilde{\mathcal{S}}_{\mathrm{D}}\right) = \operatorname{dist}\left(\operatorname{prox}_{\tilde{\delta},r}(\lambda - r\bar{s}), \tilde{\mathcal{S}}_{\mathrm{D}}\right). \tag{2.33}$$

Now, since $\tilde{\mathcal{S}}_{\mathrm{D}} = \arg\min \tilde{\delta}$ and a proximal step decreases the distance to the minimizer set (a standard property in proximality), there holds

$$\operatorname{dist}\left(\operatorname{prox}_{\tilde{\delta},r}(\lambda - r\bar{s}), \tilde{\mathcal{S}}_{\mathrm{D}}\right) \leqslant \operatorname{dist}(\lambda - r\bar{s}, \tilde{\mathcal{S}}_{\mathrm{D}}). \tag{2.34}$$

The inequality in (*iii*) is now obtained by combining (2.33), (2.34), and (*i*). □

Another way of viewing point (*ii*) is to observe that it tells us that the proximal step from λ to λ_+ on $\lambda \mapsto \delta(\lambda) = \tilde{\delta}(\lambda) + \bar{s}^{\mathsf{T}}\lambda$ is decomposed into the sum of the proximal step on the *linear* function $\lambda \to \bar{s}^{\mathsf{T}}\lambda$, from λ to $\lambda - r\bar{s}$, and the proximal step on the convex function $\tilde{\delta}$, from $\lambda - r\bar{s}$ to λ_+. The linearity of the first function is important to have that decomposition.

The next characterization of a solution to the closest feasible problem (2.24) is used in the stopping criterion of the revised version of the AL algorithm, given in section 4.2. We further discuss this matter after the proof of the proposition.

Proposition 2.18 (optimality conditions of the closest feasible problem) *Let* $r \geqslant 0$ *and let* ℓ_r *be the augmented Lagrangian* (1.5). *Then* $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times [l, u]$ *is a solution to the closest feasible problem* (2.24) *if and only if there is some* $\overline{\lambda} \in \mathbb{R}^m$ *such* *that*

$$(\overline{x}, \overline{y}) \in \underset{(x,y)\in\mathbb{R}^n\times[l,u]}{\arg\min} \quad \ell_r(x, y, \overline{\lambda}), \tag{2.35}$$

$$A^{\mathsf{T}}(A\overline{x} - \overline{y}) = 0, \tag{2.36}$$

$$P_{[l,u]}(A\overline{x}) = \overline{y}. \tag{2.37}$$

Proof. Note that in both parts of the equivalence, assumption 2.6 holds. This is clearly the case by (2.13) when the closest feasible problem has a solution. This is also the case by the implication $(iii) \Rightarrow (i)$ of proposition 2.5 and (2.12) when the augmented Lagrangian with $r > 0$ has a minimizer. Finally, it is also the case by (2.12) when the augmented Lagrangian with $r = 0$ (i.e., the Lagrangian) has a minimizer.

[Necessity] Since a solution $(\overline{x}, \overline{y})$ to the closest feasible problem (2.24) satisfies the constraints of that problem, (2.36) and (2.37) hold by the implication $(i) \Rightarrow (ii)$ of lemma 2.13. Now, by the identity $\mathcal{S} = \mathcal{R}(\partial\delta)$ in (2.15), $\bar{s} \in \mathcal{S}$ implies the existence of some $\overline{\lambda}$ such that $\bar{s} = \partial\delta(\overline{\lambda})$. By the implication $(i) \Rightarrow (iii)$ of proposition 2.9, $(\overline{x}, \overline{y})$ minimizes the Lagrangian $\ell(\cdot, \cdot, \overline{\lambda})$ on $\mathbb{R}^n \times [l, u]$:

$$q(\overline{x}) + \overline{\lambda}^{\mathsf{T}}(A\overline{x} - \overline{y}) \leqslant q(x) + \overline{\lambda}^{\mathsf{T}}(Ax - y), \qquad \forall\, (x, y) \in \mathbb{R}^n \times [l, u]. \tag{2.38}$$

Now, $y - Ax \in \mathcal{S} := [l, u] + \mathcal{R}(A)$, so that $\|A\overline{x} - \overline{y}\| = \|\bar{s}\| \leqslant \|Ax - y\|$ by the minimum norm property of $\bar{s}$ in (1.11). Using (2.38), we get for all $(x, y) \in \mathbb{R}^n \times [l, u]$:

$$q(\overline{x}) + \overline{\lambda}^{\mathsf{T}}(A\overline{x} - \overline{y}) + \frac{r}{2}\|A\overline{x} - \overline{y}\|^2 \leqslant q(x) + \overline{\lambda}^{\mathsf{T}}(Ax - y) + \frac{r}{2}\|Ax - y\|^2.$$

This is (2.35).

[Sufficiency] By the implication $(ii) \Rightarrow (i)$ of lemma 2.13, (2.36) and (2.37) show that $(\overline{x}, \overline{y})$ satisfies the constraints of the closest feasible problem (2.24). Now let (x, y) satisfy the constraints of (2.24). Then (2.35) and $A\overline{x} - \overline{y} = Ax - y = -\bar{s}$ yield

$$q(\overline{x}) - \overline{\lambda}^{\mathsf{T}}\bar{s} + \frac{r}{2}\|\bar{s}\|^2 \leqslant q(x) - \overline{\lambda}^{\mathsf{T}}\bar{s} + \frac{r}{2}\|\bar{s}\|^2.$$

Hence $q(\overline{x}) \leqslant q(x)$, implying that $(\overline{x}, \overline{\lambda})$ is a solution to (2.24). □

Since at each iteration of the AL algorithm, in step 1 actually, the condition (2.35) is satisfied with λ_k in place of $\overline{\lambda}$, it makes sense to stop the AL iterations when conditions (2.36) and (2.37) are approximately satisfied, namely when

$$A^{\mathsf{T}}(Ax_{k+1} - y_{k+1}) \simeq 0 \qquad \text{and} \qquad P_{[l,u]}(Ax_{k+1}) - y_{k+1} \simeq 0. \tag{2.39}$$

For this reason, we take these last two conditions as stopping criterion in step 3 of the revised version of the AL algorithm in section 4.2. Proposition 4.2 below will show by its points (ii) and (iii) that they are eventually satisfied by the AL algorithm.

3 Global linear convergence

With the results presented in the previous section, one can now start the analysis of the convergence of the AL algorithm when the considered QP may be infeasible. The notion of convergence will, of course, have to be redefined, since then the QP may have neither primal nor dual solution. Nevertheless, section 3.2 will show that, when assumption 2.6 holds, the AL algorithm is able to find a solution to the closest feasible problem (2.24) at a global linear speed.

Let us denote by $\{(x_k, y_k)\}$ and $\{\lambda_k\}$ the primal and dual sequences generated by the AL algorithm.

3.1 Convergence

This section deals with monotonicity and convergence properties of the AL algorithm that can be obtained without the use of an error bound of the dual solution set $\tilde{\mathcal{S}}_{\rm D}$ of the closest feasible problem (1.12). The convergence result of point (*iii*) extends a little the one by Spingarn [59; 1987, lemma 1] (see also the earlier contributions by Bruck and Reich [7, 45; 1977]), in the sense that it does not assume that the penalty parameters r_k are fixed to 1: the constraint values or dual function subgradients

$$s_k := y_k - Ax_k$$

converge to the smallest feasible shift $\bar{s}$, provided r_k is bounded away from zero. As we shall see in theorem 3.4, this convergence result prevails when the augmentation parameters r_k are small.

Proposition 3.1 (convergence without error bound) *Suppose that assumption 2.6 holds. Then*

(*i*) *the sequence* $\{\|s_k\|\}_{k\geqslant 1}$ *is nonincreasing,*

(*ii*) *the sequence* $\{\operatorname{dist}(\lambda_k, \tilde{\mathcal{S}}_{\rm D})\}_{k\geqslant 0}$ *is nonincreasing,*

(*iii*) *if* r_k *is bounded away from zero, then* $s_k \to \bar{s}$.

Proof. [(*i*)] The inequality $\|s_{k+1}\| \leqslant \|s_k\|$ is a standard property of the proximal algorithm and can be obtained by writing

$$\|s_k\|^2 = \|(s_k - s_{k+1}) + s_{k+1}\|^2 = \|s_k - s_{k+1}\|^2 + 2\langle s_k - s_{k+1}, s_{k+1}\rangle + \|s_{k+1}\|^2.$$

Observe now that the cross term in the right hand side is nonnegative by the monotonicity of $\partial\delta(\cdot)$:

$$\langle s_k - s_{k+1}, s_{k+1}\rangle = \frac{1}{r_k}\,\langle s_k - s_{k+1}, \lambda_k - \lambda_{k+1}\rangle \geqslant 0,$$

since, by (2.11), $s_k \in \partial\delta(\lambda_k)$ and $s_{k+1} \in \partial\delta(\lambda_{k+1})$. Point (*i*) follows.

[(*ii*)] Recall that $\lambda_{k+1} = \operatorname{prox}_{\delta,r_k}(\lambda_k)$ (lemma 2.4) and apply point (*iii*) of lemma 2.17.

[(iii)] The inventive idea used in [59; 1987, lemma 1] is to compare the sequence $\{\lambda_k\}_{k\geqslant 0}$ with a sequence $\{\mu_k\}_{k\geqslant 0}$ in $\tilde{\mathcal{S}}_{\rm D} \neq \varnothing$ (lemma 2.14) defined as follows:

$$\mu_0 \in \tilde{\mathcal{S}}_{\rm D} \qquad \text{and} \qquad \mu_{k+1} = \mu_k - r_k\bar{s} \quad (\text{for } k \geqslant 0).$$

By point (iii) of lemma 2.15, $\{\mu_k\} \subseteq \tilde{\mathcal{S}}_{\rm D}$. Since $\lambda_{k+1} = \lambda_k - r_k s_{k+1}$ by (1.7), there holds

$$\lambda_k - \mu_k = \lambda_{k+1} - \mu_{k+1} + r_k(s_{k+1} - \bar{s}).$$

Now $s_{k+1} \in \partial\delta(\lambda_{k+1})$, $\bar{s} \in \partial\delta(\mu_{k+1})$ by (2.30), and the monotonicity of $\partial\delta$ imply that $\langle s_{k+1} - \bar{s}, \lambda_{k+1} - \mu_{k+1}\rangle \geqslant 0$. Therefore, taking the squared norm of both sides of the identity above and neglecting the resulting cross term in the right hand side yield

$$\|\lambda_k - \mu_k\|^2 \geqslant \|\lambda_{k+1} - \mu_{k+1}\|^2 + r_k^2\|s_{k+1} - \bar{s}\|^2.$$

Since the last term is nonnegative, the inequality shows that the nonnegative sequence $\{\|\lambda_k - \mu_k\|\}$ is nonincreasing, hence converges. Therefore the same inequality implies that $r_k\|s_{k+1} - \bar{s}\|$ converges to zero. Since r_k is bounded away from zero, $s_k \to \bar{s}$. □

It will be shown in lemma 3.3, that $\operatorname{dist}(\lambda_k, \tilde{\mathcal{S}}_{\rm D})$ also tends to zero when r_k is bounded away to zero.

To extend the inequality (42) in [17], it would have been pleasant that $\|s_{k+1} - \bar{s}\|$ does not exceed $\|s_k - \bar{s}\|$, whatever is the index k and the augmentation parameter $r_k > 0$. As shown by the following example, however, it is not true that the sequence $\{\|s_k - \bar{s}\|\}_{k\geqslant 1}$ is nonincreasing for small r_k. For large r_k, section 3.2 will show that this sequence is linearly decreasing.

Example 3.2 (non monotonicity of $\{\|s_k - \bar{s}\|\}_{k\geqslant 1}$) Consider the problem (1.1), in which $n = 1$, $m = 2$, $g = 0$, $H = 1$, $A = e$ (e is the vector of all ones in $\mathbb{R}^2$), $l = (-1, 2)$, and $u = (0, 3)$. The smallest feasible shift is $\bar{s} = (-1, 1)$. Let the augmentation parameter be fixed to an arbitrary (small) constant value r in the open interval $]0, (\sqrt{2} - 1)/2[$ and let the initial iterate be $\lambda_0 = r(l_1, l_2) = r(-1, 2)$. It is easier to compute the next two iterates λ_1 and λ_2 of the AL algorithm as though they were generated by the proximal algorithm on the dual function (lemma 2.4), which reads here

$$\delta : \lambda \in \mathbb{R}^2 \mapsto \delta(\lambda) = \left(\max_{y\in[l,u]} y^{\mathsf{T}}\lambda\right) + \frac{1}{2}\,(e^{\mathsf{T}}\lambda)^2.$$

- Since $0 = \lambda_0 - r(l_1, l_2)$ and $(l_1, l_2) \in \partial\delta(0) = [l, u]$, the next iterate is $\lambda_1 = 0$, with $s_1 = (l_1, l_2) = (-1, 2)$.
- To show that λ_2, defined by $\lambda_2 := \lambda_1 - rs_2 = -rs_2$ for some $s_2 \in \partial\delta(\lambda_2)$, is the vector

  $$\lambda_2 = -\frac{r}{1 + 2r}\begin{pmatrix} -2r \\ 2 + 2r \end{pmatrix},$$

  we only have to prove that $s_2 := (-2r, 2 + 2r)/(1 + 2r)$ is in $\partial\delta(\lambda_2)$. Since the first component of λ_2 is positive and the second is negative, the expression of the dual function above shows that the function is differentiable at λ_2 and that

  $$\nabla\delta(\lambda_2) = \begin{pmatrix} u_1 \\ l_2 \end{pmatrix} + ee^{\mathsf{T}}\lambda_2 = \begin{pmatrix} 0 \\ 2 \end{pmatrix} - \frac{2r}{1 + 2r}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = s_2.$$

Now, $\|s_1 - \bar{s}\| = 1$ and $\|s_2 - \bar{s}\| = \sqrt{2}/(1+2r)$, so that the inequality $\|s_2 - \bar{s}\| \leqslant \|s_1 - \bar{s}\|$ fails for the chosen small value of r. □

3.2 Linear convergence

This section presents convergence results that prevail when the augmentation parameters r_k are sufficiently large. The results depend on the error bound associated with the dual solution set $\tilde{\mathcal{S}}_{\mathrm{D}}$, derived from the one presented in lemma 2.12.

Lemma 3.3 ($\mathrm{dist}(\boldsymbol{\lambda_k}, \tilde{\mathcal{S}}_{\mathrm{D}})$) *Suppose that assumption 2.6 holds. Then for any $\beta > 0$, there exists an $L > 0$, such that $\mathrm{dist}(\lambda_0, \tilde{\mathcal{S}}_{\mathrm{D}}) \leqslant \beta$ implies that*

$$\forall\, k \geqslant 1: \qquad \mathrm{dist}(\lambda_k, \tilde{\mathcal{S}}_{\mathrm{D}}) \leqslant L\, \|s_k - \bar{s}\|. \tag{3.1}$$

In particular, if r_k is bounded away from zero, $\mathrm{dist}(\lambda_k, \tilde{\mathcal{S}}_{\mathrm{D}}) \to 0$.

Proof. Since, by (2.14), the closest feasible problem (2.24) has a solution. The quasi-global error bound of lemma 2.12 applied to the second form of the problem in (2.24) yields

$$\begin{array}{c} \text{for any bounded set } \tilde{\mathcal{B}} \subseteq \mathbb{R}^m, \text{ there is an } L > 0, \text{ such that} \\ \forall\, \lambda \in \tilde{\mathcal{S}}_{\mathrm{D}} + \tilde{\mathcal{B}}, \quad \forall\, \tilde{s} \in \partial\tilde{\delta}(\lambda): \quad \mathrm{dist}(\lambda, \tilde{\mathcal{S}}_{\mathrm{D}}) \leqslant L\, \|\tilde{s}\|. \end{array} \tag{3.2}$$

Let $\beta > 0$ and define $\mathcal{B} := \beta B$, where B denotes the closed unit ball. Let $L > 0$ be the constant given by (3.2). Assume now that $\mathrm{dist}(\lambda_0, \tilde{\mathcal{S}}_{\mathrm{D}}) \leqslant \beta$. Then point (ii) of proposition 3.1 implies that $\mathrm{dist}(\lambda_k, \tilde{\mathcal{S}}_{\mathrm{D}}) \leqslant \beta$ for all $k \geqslant 0$, which can also be written $\lambda_k \in \tilde{\mathcal{S}}_{\mathrm{D}} + \mathcal{B}$. Assume now that $k \geqslant 1$. By (2.11) and (2.28), $s_k - \bar{s} \in \partial\tilde{\delta}(\lambda_k)$. Therefore, one can use $\lambda = \lambda_k$ and $\tilde{s} = s_k - \bar{s}$ in (3.2), which leads to (3.1).

To see that $\mathrm{dist}(\lambda_k, \tilde{\mathcal{S}}_{\mathrm{D}}) \to 0$, use point (iii) of proposition 3.1 and (3.1). □

Here is our main result.

Theorem 3.4 (global linear convergence) *Suppose that assumption 2.6 holds. Then, for any $\beta > 0$, there exists an $L > 0$, such that $\mathrm{dist}(\lambda_0, \tilde{\mathcal{S}}_{\mathrm{D}}) \leqslant \beta$ implies that*

$$\forall\, k \geqslant 1: \qquad \|s_{k+1} - \bar{s}\| \leqslant \frac{L}{r_k}\, \|s_k - \bar{s}\|, \tag{3.3}$$

$$\forall\, k \geqslant 0: \qquad \mathrm{dist}(\lambda_{k+1}, \tilde{\mathcal{S}}_{\mathrm{D}}) \leqslant \min\left(\frac{L}{r_k}, 1\right) \mathrm{dist}(\lambda_k, \tilde{\mathcal{S}}_{\mathrm{D}}). \tag{3.4}$$

Proof. Suppose that $k \geqslant 0$ and consider an arbitrary $\tilde{\lambda} \in \tilde{\mathcal{S}}_{\mathrm{D}}$. First $0 \in \partial\tilde{\delta}(\tilde{\lambda})$, by the definition (2.29) of $\tilde{\mathcal{S}}_{\mathrm{D}}$. Next $s_{k+1} - \bar{s} \in \partial\delta(\lambda_{k+1}) - \bar{s} = \partial\tilde{\delta}(\lambda_{k+1})$, by (2.11) and (2.28). Then, the monotonicity of the multifunction $\partial\tilde{\delta}$ implies that

$$(s_{k+1} - \bar{s})^{\mathsf{T}}(\lambda_{k+1} - \tilde{\lambda}) \geqslant 0. \tag{3.5}$$

On the other hand, subtracting $\tilde{\lambda}+r_k\bar{s}$ from both sides of the iteration identity (1.17) and introducing $\tilde{\lambda}_k := \lambda_k - r_k\bar{s}$ yield

$$\lambda_{k+1}-\tilde{\lambda}+r_k(s_{k+1}-\bar{s}) = \tilde{\lambda}_k-\tilde{\lambda}.$$

Taking the squared norm of both sides of this identity, using (3.5) and $r_k>0$, and neglecting $\|\lambda_{k+1}-\tilde{\lambda}\|^2$ lead to

$$\|s_{k+1}-\bar{s}\| \leqslant \frac{1}{r_k}\|\tilde{\lambda}_k-\tilde{\lambda}\|.$$

Since $\tilde{\lambda}$ is arbitrary in $\tilde{\mathcal{S}}_{\text{D}}$:

$$\|s_{k+1}-\bar{s}\| \leqslant \frac{1}{r_k}\operatorname{dist}(\tilde{\lambda}_k,\tilde{\mathcal{S}}_{\text{D}}).$$

Now the expression of $\tilde{\lambda}_k = \lambda_k - r_k\bar{s}$ and point (i) of lemma 2.17 yield

$$\forall\, k\geqslant 0: \qquad \|s_{k+1}-\bar{s}\| \leqslant \frac{1}{r_k}\operatorname{dist}(\lambda_k,\tilde{\mathcal{S}}_{\text{D}}). \tag{3.6}$$

Assuming that $k\geqslant 1$ and using (3.1) in (3.6) gives (3.3).

On the other hand, starting with (3.1) and using (3.6) lead to (3.4) with the factor L/r_k. For getting the unit factor in (3.4), just use point (ii) of proposition 3.1. □

4 The revised AL algorithm

4.1 Update of the augmentation parameters

When the convex quadratic optimization problem (1.1) has a solution, the estimate (1.8) offers a possibility to design an update rule for the augmentation parameters r_k, based on a desired linear convergence rate $\rho_{\text{des}} \in \,]0,1[$ of the constraint value $s_k := y_k - Ax_k$ towards zero (the lower ρ_{des} is, the faster the convergence is required). In practice, this convergence rate is easier to specify by the user of the algorithm than the augmentation parameter itself, because a satisfactory value of the latter depends in a complex way on the problem data and its solutions (see formula (34) in [17]). The rule proposed in [17] and implemented in [27, 28] is based on an examination of the ratio $\rho_k := \|s_{k+1}\|/\|s_k\|$: if this one is not less than ρ_{des}, r_{k+1} is set to $r_k\rho_k/\rho_{\text{des}}$. The logic is that, from (1.8), ρ_k is always less than L/r_k, so that it makes sense to increase r_k in this way.

When problem (1.1) is infeasible, the ratio $\|s_{k+1}\|/\|s_k\|$ is no longer bounded by L/r_k (it cannot be, since s_k cannot tend to zero), so that the update rule of r_k sketched above generates an unbounded sequence of augmentation parameters, without the hope to realize what it is designed for. By theorem 3.4, $\|s_{k+1}-\bar{s}\|/\|s_k-\bar{s}\|$ is bounded by L/r_k, but the latter ratio is not accessible while the algorithm is running, since the smallest feasible shift $\bar{s}$ is not known before convergence is reached, so that the extension of the above update rule to infeasible problems is not straightforward. In other to overcome this difficulty, we propose to watch instead the ratio $\|s'_{k+1}\|/\|s'_k\|$, where

$$s'_k = s_k - s_{k-1}. \tag{4.1}$$

This proposal is grounded on the following proposition.

Proposition 4.1 *Let $\{s_k\}$ and $\{s'_k\}$ be two sequences of a normed space $\mathbb{E}$, whose elements are linked by (4.1).*

1) *If the sequence $\{s_k\}$ satisfies for some $\bar{s} \in \mathbb{E}$, some $\rho \in [0, 1[$, and some index k_1*

$$\forall\, k \geqslant k_1 : \quad \|s_{k+1} - \bar{s}\| \leqslant \rho \|s_k - \bar{s}\|,$$

then the sequence $\{s'_k\}$ verifies

$$\forall\, k \geqslant k_1 + 1 : \quad \|s'_{k+1}\| \leqslant \frac{(1+\rho)\rho}{1-\rho} \|s'_k\|. \tag{4.2}$$

2) *Conversely, if the sequence $\{s'_k\}$ verifies for some $\rho' \in [0, 1[$ and some index k_1*

$$\forall\, k \geqslant k_1 : \quad \|s'_{k+1}\| \leqslant \rho' \|s'_k\|,$$

then the sequence $\{s_k\}$ converges to some $\bar{s}$ and satisfies

$$\forall\, k \geqslant k_1 - 1 : \quad \|s_{k+1} - \bar{s}\| \leqslant \frac{\rho'}{1-2\rho'} \|s_k - \bar{s}\|. \tag{4.3}$$

Proof. 1) Let $k \geqslant k_1 + 1$. Then

$$\begin{aligned} \|s'_{k+1}\| &\leqslant \|s_{k+1} - \bar{s}\| + \|s_k - \bar{s}\| \leqslant (1+\rho)\|s_k - \bar{s}\| \\ \|s'_k\| &\geqslant \|s_{k-1} - \bar{s}\| - \|s_k - \bar{s}\| \geqslant (1-\rho)\|s_{k-1} - \bar{s}\|. \end{aligned}$$

Hence

$$\|s'_{k+1}\| \leqslant (1+\rho)\|s_k - \bar{s}\| \leqslant (1+\rho)\rho\|s_{k-1} - \bar{s}\| \leqslant \frac{(1+\rho)\rho}{1-\rho}\|s'_k\|.$$

2) Observe first that the sequence $\{s_k\}$ is a Cauchy sequence, since for $l > k \geqslant k_1 - 1$, there holds

$$\|s_l - s_k\| \leqslant \|s'_l\| + \cdots + \|s'_{k+1}\| \leqslant \sum_{i=0}^{l-k-1} (\rho')^i \|s'_{k+1}\| \leqslant \frac{1}{1-\rho'}\|s'_{k+1}\|,$$

which tends to zero when $k \to \infty$. Therefore $\{s_k\}$ converges, say to some $\bar{s}$. Taking $k+1$ instead of k in the previous estimate and letting $l \to \infty$ yield for $k \geqslant k_1 - 1$:

$$\|s_{k+1} - \bar{s}\| \leqslant \frac{1}{1-\rho'}\|s'_{k+2}\| \leqslant \frac{\rho'}{1-\rho'}\|s'_{k+1}\| \leqslant \frac{\rho'}{1-\rho'}(\|s_{k+1} - \bar{s}\| + \|s_k - \bar{s}\|).$$

The linear convergence of $\{s_k\}$ in (4.3) follows. $\square$

This proposition shows that the linear convergence of the sequence $\{s_k - \bar{s}\}$ and $\{s'_k\}$ occurs simultaneously, provided their convergence rate is sufficiently small:

- by point 1: as soon as the sequence $\{s_k\}$ converges linearly to some $\bar{s}$ with a rate $\rho < \sqrt{2}-1 \simeq 0.41$, the sequence $\{s'_k\}$ converges linearly to zero with the rate $(1+\rho)\rho/(1-\rho)$;
- by point 2: as soon as the sequence $\{s'_k\}$ converges linearly to zero with a rate $\rho' < 1/3$, the sequence $\{s_k\}$ converges linearly to some $\bar{s}$ with the rate $\rho'/(1-2\rho')$.

For example, if $\rho = 0.1$, the rate of convergence in (4.2) is approximately 0.122; if $\rho' = 0.1$ the rate of convergence in (4.3) is 0.125.

4.2 Revised augmented Lagrangian algorithm

Let us now incorporate in the AL algorithm of section 1 the modifications suggested by the analysis of this paper: a new stopping criterion is introduced in step 3 and a new rule for updating the augmentation parameter is found in step 4. The algorithm is described as though computation were done in exact arithmetic.

Revised AL algorithm to solve (1.1)

Initialization: choose $\lambda_0 \in \mathbb{R}^m$, $r_0 > 0$, and $\rho_{\rm des} \in \,]0,1[$; set $\rho'_{\rm des} := \rho_{\rm des}/(1+2\rho_{\rm des})$.
Repeat for $k = 0, 1, 2, \ldots$

1. If the feasible AL subproblem (1.15) has no solution, exit with a direction $d \in \mathbb{R}^n$ verifying (1.16). Otherwise, denote a solution to (1.15) by (x_{k+1}, y_{k+1}).
2. Update the multiplier by (1.17).
3. Stop if

$$A^{\mathsf{T}}(Ax_{k+1} - y_{k+1}) = 0 \qquad \text{and} \qquad P_{[l,u]}(Ax_{k+1}) = y_{k+1}.$$

4. Update the augmentation parameter if $k \geqslant 1$. Let s_{k+1} and s'_{k+1} be given by (1.13) and (4.1) respectively and set $\rho'_k := \|s'_{k+1}\|/\|s'_k\|$. Then take

$$r_{k+1} := \max\left(1, \frac{\rho'_k}{\rho'_{\rm des}}\right) r_k. \tag{4.4}$$

The next paragraphs discuss the new components of the algorithm.

By proposition 2.5, if assumption 2.6 does not hold, the revised AL algorithm exits in step 1 at the first iteration ($k = 0$). Otherwise, the stopping criterion in step 3 is eventually satisfied (up to a given precision), as shown by the next proposition.

Proposition 4.2 (satisfaction of the stopping criterion) *Suppose that assumption 2.6 holds. Then the revised AL algorithm does not terminate in step 1 with a direction of unboundedness and generates a sequence $\{(x_k, y_k)\}$ that satisfies*

$$A^{\mathsf{T}}(Ax_k - y_k) \to 0 \qquad \textit{and} \qquad P_{[l,u]}(Ax_k) - y_k \to 0. \tag{4.5}$$

PROOF. By assumption 2.6, the AL subproblems have a solution, so that the algorithm does not terminate in step 1 with a direction of unboundedness.

By proposition 3.1 and the fact that r_k is bounded away from zero (it can only increase in this version of the algorithm), $s_k := y_k - Ax_k \to \bar{s}$, so that $A^{\mathsf{T}}(y_k - Ax_k) \to A^{\mathsf{T}}\bar{s} = 0$ by (2.22), which is the first condition in (4.5).

Let us denote the projection of Ax_k on $[l, u]$ by

$$\tilde{y}_k := P_{[l,u]}(Ax_k),$$

which is characterized by

$$(\tilde{y}_k - Ax_k)^{\mathsf{T}}(y - \tilde{y}_k) \geqslant 0, \qquad \forall\, y \in [l, u].$$

Taking $y = y_k \in [l, u]$ yields

$$(\tilde{y}_k - Ax_k)^{\mathsf{T}}(y_k - \tilde{y}_k) \geqslant 0. \tag{4.6}$$

On the other hand, the characterization of the projection $\bar{s}$ of zero on $\mathcal{S}$ can be written

$$\bar{s}^{\mathsf{T}}(s - \bar{s}) \geqslant 0, \qquad \forall\, s \in \mathcal{S}.$$

Taking $s = \tilde{y}_k - y_k + s_k = \tilde{y}_k - Ax_k \in [l, u] + \mathcal{R}(A) = \mathcal{S}$ yields

$$\bar{s}^{\mathsf{T}}(\tilde{y}_k - y_k + s_k - \bar{s}) \geqslant 0. \tag{4.7}$$

Now adding (4.6) and (4.7) leads to

$$(\bar{s} - \tilde{y}_k + Ax_k)^{\mathsf{T}}(\tilde{y}_k - y_k) + \bar{s}^{\mathsf{T}}(s_k - \bar{s}) \geqslant 0$$

or, using $s_k := y_k - Ax_k$ and Cauchy-Schwarz inequality,

$$\|\tilde{y}_k - y_k\|^2 \leqslant (\bar{s} - s_k)^{\mathsf{T}}(\tilde{y}_k - y_k) + \bar{s}^{\mathsf{T}}(s_k - \bar{s}) \leqslant \|s_k - \bar{s}\|\,\|\tilde{y}_k - y_k\| + \|\bar{s}\|\,\|s_k - \bar{s}\|.$$

This inequality, quadratic in $\|\tilde{y}_k - y_k\|$, and the convergence $s_k \to \bar{s}$ imply that there is a constant $\gamma > 0$ such that

$$\|\tilde{y}_k - y_k\| \leqslant \gamma \|s_k - \bar{s}\|^{1/2},$$

which certainly implies the second condition in (4.5). $\square$

The logic behind the update rule of the augmentation parameter r_k in step 4 is the following. The algorithm should ideally guarantee the desired convergence rate $\rho_{\rm des} \in]0,1[$ of $s_k := y_k - Ax_k$ towards $\bar{s}$, as opposed to the one of $\{s'_k\}$ to zero, because this convergence, expressed in terms of the optimization problem data, is meaningful for the user of the algorithm. Nevertheless, we have already pointed out that the current value of the quotient

$$\rho_k := \frac{\|s_{k+1} - \bar{s}\|}{\|s_k - \bar{s}\|}$$

cannot be examined ($\bar{s}$ being unknown), so that the algorithm tries to get the convergence rate of $\rho'_{\rm des} := \rho_{\rm des}/(1 + 2\rho_{\rm des})$ on s'_k, which implies indeed a rate $\rho_{\rm des}$ for the linear convergence of s_k towards $\bar{s}$ (see point 2 of proposition 4.1). Now, if the effect of r_k on the rate of convergence of s_k to $\bar{s}$ is transparent through (3.3), its effect on the rate of convergence of $\{s'_k\}$ is more complex. For this reason, if we assume that $\rho_{\rm des}$ is sufficiently small, say less than 0.1, the rate of convergence of the two sequences $\{s_k - \bar{s}\}$ and $\{s'_k\}$ are close to each other (proposition 4.1), and the algorithm can proceed on $\{s'_k\}$ as it would do on $\{s_k\}$: if

$$\rho'_k := \frac{\|s'_{k+1}\|}{\|s'_k\|}$$

is sufficiently small (step 4.1), the value of r_k is unchanged; otherwise, r_k is multiplied by the factor $\rho'_k/\rho_{\rm des}$ (step 4.2).

The update rule of the augmentation parameters in step 4 will maintain the sequence $\{r_k\}$ bounded, even when the quadratic problem is infeasible, since $\rho'_k \leqslant \rho'_{\rm des}$ as soon as the quotient ρ_k is permanently less than the positive root $\rho^+_{\rm des}$ of $\rho \mapsto \rho^2 + (1 + \rho'_{\rm des})\rho - \rho'_{\rm des}$ (point 1 of proposition 4.1), which will occur if r_k is permanently larger than $L/\rho^+_{\rm des}$ (inequality (3.3)). As already observe by Fortin and Glowinski [24; 1982, remark 5.6, page 42], if the generated multipliers λ_k blow up when the problem is infeasible, they do so by adding at each iteration the converging term $-r_k s_k \to -r\bar{s}$ (if r_k converges to r), which is much slower than the decrease of $s_k \to 0$, which occurs with a linear convergence speed. Hence overflow will not be observed in the implementation of the algorithm.

5 Perspectives

The implementation of the revised AL algorithm of section 4.2 for solving convex quadratic optimization problems is ongoing. It takes the form of `C++`/`Matlab` pieces of software called `Oqla`/`Qpalm` [28]. A particular attention is paid to the problems that are either unbounded or infeasible. When the closest feasible problem is unbounded, the codes return an *unboundedness direction*, that is a direction d satisfying (2.4). Otherwise, assumption 2.6 holds and the codes return a solution to the closest feasible problem (1.12), as well as the smallest feasible shift $\bar{s}$ (which can vanish). These features are attractive when the solvers are used to deal with the convex quadratic optimization problems generated by some versions of the SQP algorithm for solving nonlinear optimization problems.

It would be interesting to know whether the global linear convergence presented in this paper can be extended to a (possibly infeasible) convex quadratic problem defined on a

Euclidean space $\mathbb{E}$ (with a scalar product denoted by $\langle\cdot,\cdot\rangle$) and that reads

$$\left\{\begin{array}{l} \inf_x \langle g,x\rangle + \frac{1}{2}\langle x,Hx\rangle \\ Ax \in C \\ x \in X, \end{array}\right. \tag{5.1}$$

where $g \in \mathbb{E}$, H is a linear symmetric positive definite operator on $\mathbb{E}$, A is a linear operator from $\mathbb{E}$ to some linear space $\mathbb{F}$, while X and C are convex sets in $\mathbb{E}$ and $\mathbb{F}$ respectively. The experience acquired in [17] and in this paper suggests that the polyhedrality of the sets X and C is probably sufficient to get the quasi-global error bound of lemma 2.12, which has been important so far to get the global linear convergence result, but other assumptions on X and C might also yield a similar error bound. The generalization (5.1) of (1.1) is useful, in particular, because it can model by $x \in X$ a trust region constraint [14], which is not present in (1.1) as a constraint satisfied at each iteration of the AL algorithm, even when the constraints are incompatible. Such a constraint may occur in a trust region approach for solving a nonlinear optimization problem or may be used to prevent the solution to (1.1) from being discontinuous with respect to the problem data [10, 8, 9]. Many algorithms have indeed been proposed to find a solution to some relaxed version of (5.1) when the set $\{x \in \mathbb{E} : Ax \in C,\, x \in X\}$ is empty (see [43, 60, 13, 12, 40, 44] to mention a few), while to our knowledge the use of the AL algorithm has not been investigated.

Another computationally important question is to know whether the linear convergence result still holds when the AL subproblems are solved inexactly. Contributions along this line include [31, 55, 56, 57, 22, 23] and the references thereof.

The case of the Lagrangian relaxation algorithm [34; chapters XIV-XV] probably deserves more investigations. Indeed, on the one hand, Dean and Glowinski [16; 2006, theorem 4.1] have shown that, for the minimization of a strictly convex quadratic function subject to linear equality constraints, the Lagrangian relaxation method, which is the steepest descent algorithm on the dual function, with sufficiently small step-sizes in the dual space, generates primal iterates that converge globally linearly to the (unique) solution to the closest feasible problem. On the other hand, more robust and accurate algorithms like bundle methods have been shown to be possible approaches to computing an approximate proximal point [36, 15, 4; 1990-1995], so that their use on the dual function could benefit from the properties of the augmented Lagrangian (or proximal method on the dual function) highlighted in this paper. We are not aware of an extension of this result to convex problems with inequality constraints.

Acknowledgment

We would like to thank Arnaud Lenoir for discussions on the proximal algorithm, in particular for having drawn our attention on the contributions by Bruck and Reich [7, 45; 1977] and Spingarn [59; 1987], which helped us to make progress on the subject even though these results do not intervene in the present developments (Spingarn's technique for proving $s_k \to \bar{s}$ is however mentioned in the proof of proposition 3.1, for completeness). We are also most grateful to Claude Lemaréchal for discussions that result in the second part of the last paragraph of the *Perspectives* section. We also thank a referee for having quoted the partial link between proposition 2.8 and theorem 11.42 and example 11.43 in [53].

The work was partially supported by the contract CIFRE number 635/2008 of the ANRT (Association Nationale de la Recherche et de la Technologie, France).

References

[1] K.J. Arrow, F.J. Gould, S.M. Howe (1973). A generalized saddle point result for constrained optimization. *Mathematical Programming*, 5, 225–234. [doi]. 7

[2] A. Auslender, M. Teboulle (2003). *Asymptotic Cones and Functions in Optimization and Variational Inequalitites*. Springer Monographs in Mathematics. Springer, New York. 8

[3] E.G. Birgin, J.M. Martínez (2014). *Practical Augmented Lagrangian Methods for Constrained Optimization*. SIAM Publication, Philadelphia. [doi]. 3

[4] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, C. Sagastizábal (1995). A family of variable metric proximal methods. *Mathematical Programming*, 68, 15–47. [doi]. 30

[5] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, C. Sagastizábal (2006). *Numerical Optimization – Theoretical and Practical Aspects* (second edition). Universitext. Springer Verlag, Berlin. 6

[6] J.M. Borwein, A.S. Lewis (2000). *Convex Analysis and Nonlinear Optimization*. Springer, New York. 5

[7] R.E. Bruck, S. Reich (1977). Nonexpansive projections and resolvents of accretive operators in Banach spaces. *Houston Journal of Mathematics*, 3, 459–470. 3, 22, 30

[8] J.V. Burke (1989). A sequential quadratic programming method for potentially infeasible mathematical programs. *Journal of Mathematical Analysis and Applications*, 139, 319–351. [doi]. 30

[9] J.V. Burke (1992). A robust trust region method for constrained nonlinear programming problems. *SIAM Journal on Optimization*, 2, 325–347. [doi]. 30

[10] J.V. Burke, S.-P. Han (1989). A robust sequential quadratic programming method. *Mathematical Programming*, 43, 277–303. [doi]. 30

[11] J.D. Buys (1972). *Dual algorithms for constrained optimization*. PhD Thesis, Rijksuniversiteit te Leiden, Leiden, The Netherlands. 7

[12] R.H. Byrd (1987, May). Robust trust region methods for constrained optimization. Third SIAM Conference on Optimization, Houston, TX. 30

[13] R.H. Byrd, R.B. Schnabel, G.A. Shultz (1987). A trust region algorithm for nonlinearly constrained optimization. *SIAM Journal on Numerical Analysis*, 24, 1152–1170. [doi]. 30

[14] A.R. Conn, N.I.M. Gould, Ph.L. Toint (2000). *Trust-Region Methods*. MPS-SIAM Series on Optimization 1. SIAM and MPS, Philadelphia. [doi]. 30

[15] R. Correa, C. Lemaréchal (1993). Convergence of some algorithms for convex minimization. *Mathematical Programming*, 62, 261–275. [doi]. 30

[16] E.J. Dean, R. Glowinski (2006). An augmented Lagrangian approach to the numerical solution of the Dirichlet problem for the elliptic Monge-Ampère equation in two dimensions. *Electronic Transactions on Numerical Analysis*, 22, 71–96. [pdf]. 4, 7, 30

[17] F. Delbos, J.Ch. Gilbert (2005). Global linear convergence of an augmented Lagrangian algorithm for solving convex quadratic optimization problems. *Journal of Convex Analysis*, 12, 45–69. [journal]. 2, 3, 4, 6, 7, 10, 16, 23, 25, 30

[18] Z. Dostál, A. Friedlander, S.A. Santos (1999). Augmented Lagrangians with adaptive precision control for quadratic programming with equality constraints. *Computational Optimization and Applications*, 14, 37–53. [doi]. 7

[19] Z. Dostál, A. Friedlander, S.A. Santos (2003). Augmented Lagrangians with adaptive precision control for quadratic programming with simple bounds and equality constraints. *SIAM Journal on Optimization*, 13, 1120–1140. [doi]. 7

[20] Z. Dostál, A. Friedlander, S.A. Santos, K. Alesawi (2000). Augmented Lagrangians with adaptive precision control for quadratic programming with equality constraints: corrigendum and addendum. *Computational Optimization and Applications*, 23, 127–133. [doi]. 7

[21] J. Eckstein (2012). Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results. RRR 32-2012, Department of Management Science and Information Systems (MSIS) and RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway, NJ 08854-8003. [preprint]. 6

[22] J. Eckstein, P.J.S. Silva (2010). Proximal methods for nonlinear programming: double regularization and inexact subproblems. *Computational Optimization and Applications*, 46, 279–304. [doi]. 30

[23] J. Eckstein, P.J.S. Silva (2013). A practical relative error criterion for augmented Lagrangians. *Mathematical Programming*, 141, 319–348. [doi]. 30

[24] M. Fortin, R. Glowinski (1982). *Méthodes de Lagrangien Augmenté – Applications à la Résolution Numérique de Problèmes aux Limites*. Méthodes Mathématiques de l'Informatique 9. Dunod, Paris. 4, 29

[25] M. Frank, P. Wolfe (1956). An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3, 95–110. [doi]. 2, 7, 8

[26] M. Friedlander, S. Leyffer (2008). Global and finite termination of a two-phase augmented Lagrangian filter method for general quadratic programs. *SIAM Journal on Scientific Computing*, 30, 1706–1726. [doi]. 7

[27] J.Ch. Gilbert (2009). QPAL – A solver of convex quadratic optimization problems, using an augmented Lagrangian approach – Version 0.6.1. Rapport Technique 0377, INRIA, BP 105, 78153 Le Chesnay, France. [pdf]. 3, 25

[28] J.Ch. Gilbert, É. Joannopoulos (2014). OQLA/QPALM - Convex quadratic optimization solvers using the augmented Lagrangian approach, able to deal with infeasibility. Technical report, INRIA, BP 105, 78153 Le Chesnay, France. To appear. 25, 29

[29] P.E. Gill, E. Wong (2012). Sequential quadratic programming methods. In J. Lee, S. Leyffer (editors), *Mixed Integer Nonlinear Programming*, volume 154 of *The IMA Volumes in Mathematics and its Applications*, pages 147–224. Springer. 6

[30] R. Glowinski, P. Le Tallec (1989). *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. SIAM Studies in Applied Mathematics 9. SIAM, Philadelphia, PA, USA. [doi]. 4

[31] O. Güler (1992). Augmented Lagrangian algorithms for linear programming. *Journal of Optimization Theory and Applications*, 75, 445–470. [doi]. 7, 30

[32] D. Henrion, J. Malick (2009). Projection methods for conic feasibility problems: applications to polynomial sum-of-squares decompositions. *Optimization Methods and Software*, 26, 23–46. [doi]. 7

[33] M.R. Hestenes (1969). Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4, 303–320. [doi]. 7

[34] J.-B. Hiriart-Urruty, C. Lemaréchal (1993). *Convex Analysis and Minimization Algorithms*. Grundlehren der mathematischen Wissenschaften 305-306. Springer-Verlag. 5, 9, 30

[35] A.F. Izmailov, M.V. Solodov (2014). *Newton-Type Methods for Optimization and Variational Problems*. Springer Series in Operations Research and Financial Engineering. Springer. 6

[36] K.C. Kiwiel (1990). Proximity control in bundle methods for convex nondifferentiable minimization. *Mathematical Programming*, 46, 105–122. [doi]. 30

[37] J. Malick (2004). A dual approach to semidefinite least-squares problems. *SIAM Journal on Matrix Analysis and Applications*, 26, 272–284. [doi]. 7

[38] J. Malick, J. Povh, F. Rendl, A. Wiegele (2009). Regularization methods for semidefinite programming. *SIAM Journal on Optimization*, 20, 336–356. [doi]. 7

[39] J.J. Moreau (1965). Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93, 273–299. [url]. 9

[40] E.O. Omojokun (1991). *Trust region algorithms for optimization with nonlinear equality and inequality constraints*. PhD Thesis, Department of Computer Science, University of Colorado, Boulder, Colorado 80309. 30

[41] B.T. Poljak, N.V. Tret'jakov (1972). A certain iteration method of linear programming and its economic interpretation (in Russian). *Èkonomika i Matematicheskie Metody*, 8, 740–751. 7

[42] M.J.D. Powell (1969). A method for nonlinear constraints in minimization problems. In R. Fletcher (editor), *Optimization*, pages 283–298. Academic Press, London. 7

[43] M.J.D. Powell (1978). A fast algorithm for nonlinearly constrained optimization calculations. In G.A. Watson (editor), *Numerical Analysis Dundee 1977*, Lecture Notes in Mathematics 630, pages 144–157. Springer-Verlag, Berlin. 30

[44] M.J.D. Powell, Y. Yuan (1991). A trust region algorithm for equality constrained optimization. *Mathematical Programming*, 49, 189–211. [doi]. 30

[45] S. Reich (1977). On infinite products of resolvents. *Rend. Classe Sci. Fis. Mat. e Nat. Accad. Naz. Lincei Ser. VIII*, LXIII, Fasc. 5. 3, 22, 30

[46] R.T. Rockafellar (1970). *Convex Analysis*. Princeton Mathematics Ser. 28. Princeton University Press, Princeton, New Jersey. 5, 13, 14, 15

[47] R.T. Rockafellar (1971). New applications of duality in convex programming. In *Proceedings of the 4th Conference of Probability, Brasov, Romania*, pages 73–81. 7

[48] R.T. Rockafellar (1973). A dual approach to solving nonlinear programming problems by unconstrained optimization. *Mathematical Programming*, 5, 354–373. [doi]. 3, 9

[49] R.T. Rockafellar (1973). The multiplier method of Hestenes and Powell applied to convex programming. *Journal of Optimization Theory and Applications*, 12, 555–562. [doi]. 7

[50] R.T. Rockafellar (1974). Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control*, 12, 268–285. 7

[51] R.T. Rockafellar (1974). *Conjugate Duality and Optimization*. Regional Conference Series in Applied Mathematics 16. SIAM, Philadelphia, PA, USA. 13

[52] R.T. Rockafellar (1976). Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14, 877–898. [doi]. 3

[53] R.T. Rockafellar, R. Wets (1998). *Variational Analysis*. Grundlehren der mathematischen Wissenschaften 317. Springer. 13, 31

[54] A. Shapiro, J. Sun (2004). Some properties of the augmented Lagrangian in cone constrained optimization. *Mathematics of Operations Research*, 29, 479–491. [doi]. 7

[55] M.V. Solodov, B.F. Svaiter (1999). A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis*, 7, 323–345. [doi]. 30

[56] M.V. Solodov, B.F. Svaiter (1999). A hybrid projection-proximal point algorithm. *Journal of Convex Analysis*, 6, 59–70. [journal]. 30

[57] M.V. Solodov, B.F. Svaiter (2000). An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Mathematics of Operations Research*, 25, 214–230. [doi]. 30

[58] J.E. Spingarn (1983). Partial inverse of a monotone operator. *Applied Mathematics and Optimization*, 10, 247–265. [doi]. 3

[59] J.E. Spingarn (1987). A projection method for least-squares solutions to overdetermined systems of linear inequalities. *Linear Algebra and its Applications*, 86, 211–236. [doi]. 3, 22, 23, 30

[60] A. Vardi (1985). A trust region algorithm for equality constrained minimization: convergence properties and implementation. *SIAM Journal on Numerical Analysis*, 22, 575–591. [doi]. 30

[61] Z. Wen, D. Goldfarb, W. Yin (2010). Alternating direction augmented Lagrangian methods for semidefinite programming. *Mathematical Programming Computation*, 2, 203–230. [doi]. 7