Dual convergence for penalty proximal point algorithms in convex programming

Felipe Alvarez[*], Miguel Carrasco[†] & Thierry Champion[‡]

Abstract

We consider an implicit iterative method in convex programming which combines inexact variants of the proximal point algorithm, with parametric penalty functions. We investigate a multiplier sequence which is explicitly computed in terms of the primal sequence generated by the iterative method, providing some conditions on the parameters in order to ensure convergence towards a particular dual optimal solution.

Keywords. Convex programming, prox method, penalty schemes.

1 Introduction

Let us consider the mathematical programming problem

$$(P) \qquad \min \{ f_0(x) \mid f_i(x) \le 0 \ \forall\, i = 1, \dots, m \}$$

where $f_i : \mathbb{R}^d \to \mathbb{R}$ is convex and differentiable for every $i \in \{0, \dots, m\}$. We assume that the optimal solution set $S(P)$ is nonempty and bounded. In order to solve (P) approximately, we take

$$f_r(x) := f_0(x) + r \sum_{i=1}^{m} \theta(f_i(x)/r),$$

where $r > 0$ is a real parameter and the penalty function $\theta(\cdot)$ satisfies the following conditions:

- $\theta :]-\infty, \kappa[\to \mathbb{R}$ is strictly convex and smooth, where $\kappa \ge 0$,
- $\theta'(u) > 0$, $\theta'(u) \to 0$ as $u \to -\infty$ and $\theta'(u) \to +\infty$ as $u \nearrow \kappa$.

For instance, let us take the exponential penalty $\theta_1(u) = \exp(u)$ with $\kappa = +\infty$, or the log-barrier $\theta_2(u) = -\ln(-u)$ and the inverse barrier $\theta_3(u) = -1/u$ both with $\kappa = 0$. When $\kappa = 0$ we assume that the standard *Slater condition* holds:

$$\exists x_0 \in \mathbb{R}^d,\ \forall i \in \{1, \dots, m\}, \qquad f_i(x_0) < 0. \tag{1}$$

[*]Centro de Modelamiento Matemático (CNRS UMI 2807), Departamento de Ingeniería Matemática, Universidad de Chile, Av. Blanco Encalada 2120, Santiago, Chile.

[†]Departamento de Ingeniería, Universidad de los Andes, Av. San Carlos de Apoquindo 2200, Las Condes, Chile.

[‡]Laboratoire Imath, U.F.R. des Sciences et Techniques, Université du Sud Toulon-Var, Avenue de l'Université, BP 20132, 83957 La Garde cedex, France; & Centro de Modelamiento Matemático (CNRS UMI 2807), Universidad de Chile, Av. Blanco Encalada 2120, Santiago, Chile.

Notice that $f_r(\cdot)$ is convex and differentiable. In addition, letting $r \to 0^+$ for a given $x \in \mathbb{R}^d$ we get $f_r(x) \to f_0(x)$ when $f_i(x) \le 0$ for all $i \in \{1, \dots, m\}$, otherwise $f_r(x) \to +\infty$. Under fairly general conditions, for each $r > 0$, there exists a unique optimal solution $x(r)$ for the penalized problem

$$(P_r) \qquad \min_{x \in \mathbb{R}^d} f_r(x).$$

Moreover, $x(r)$ converges to some $x^\theta \in S(P)$ as $r \to 0^+$. For further details on the asymptotic analysis of such *primal optimal paths*, the reader is referred to [7, 12, 14, 15].

In practice, the exact computation of $x(r)$ is replaced with inexact iterative schemes for finding an approximation x^k of $x(r_k)$, for some sequence (r_k) with $r_k > 0$ and $r_k \to 0$ as $k \to \infty$. Such "path-following" procedures may be quite expensive in computational terms if we want to ensure that x^k be close enough to $x(r_k)$, as this may degrade the numerical performance of the algorithm. Since the goal is not (P_{r_k}) but (P), an alternative approach consists in obtaining x^k by performing a prescribed number of iterations of an optimization method applied to f_{r_k}, and then update the penalty parameter to $r_{k+1} < r_k$.

In this paper, we focus on a special class of iterative methods for solving (P), in which the *proximal point algorithm* is coupled with the penalty function approach. More precisely, we consider the sequences $(x^k)_k$ generated by the following inexact implicit iterative scheme: given the current iterate x^{k-1}, a step-size $h_k > 0$ and a tolerance parameter $\varepsilon_k \ge 0$, we find x^k and g^k such that

$$x^k = x^{k-1} - h_k g^k + \eta^k, \quad g^k \in \partial_{\varepsilon_k} f_{r_k}(x^k + e^k), \qquad \text{(Penalty-PPA)}$$

for some errors e^k, $\eta^k \in \mathbb{R}^d$ which are intended to be small. Here $\partial_\varepsilon f_r(x)$ stands for the approximate ε-subdifferential of f_r at x. When $r_k \to 0$, and under appropriate conditions, (Penalty-PPA) generates sequences $(x^k)_k$ such that $x^k \to x^\infty$ as $k \to +\infty$ for some $x^\infty \in S(P)$; it may happen that $x^\infty \neq x^\theta$ in the case of multiple optimal solutions. See [4, 5, 13, 15] for results in this direction. A great effort has been devoted to relax the conditions on $(r_k)_k$ in order to ensure primal convergence. This is motivated by the practical implementation of some algorithms where the penalty parameters are updated by an adaptable *feedback* rule depending on the current iterate; an idea which may be useful to accelerate the convergence of the algorithm by avoiding numerical instabilities.

Following previous works in the linear case [5, 15] (see also §3), we associate with (x^k) a *multiplier sequence* $(\lambda^k) \subset \mathbb{R}^m_{++}$ explicitly given by

$$\lambda_i^k := \theta' \left[f_i(x^k + e^k)/r_k \right], \quad i = 1, ..., m. \qquad (2)$$

The question is under which conditions, especially on the parameters of the algorithm, we can ensure the convergence of (λ^k) to a dual optimal solution. Since we already know that $x^k \to x^\infty \in S(P)$ as $k \to \infty$, we deduce that $\lambda_i^k \to 0$ for any i such that $f_i(x^\infty) < 0$. The delicate problem is indeed the convergence for $i \in I_\infty := \{i \mid f_i(x^\infty) = 0\}$.

Under compactness of the dual optimal set, it is possible to verify that $(\lambda^k)_k$ is bounded and every cluster point is indeed a dual optimal solution. It is apparent that convergence towards a unique dual optimal solution is much harder to ensure. In this paper we give a general convergence result under minimal assumptions on the penalty parameter sequence $(r_k)_k$.

Let us mention that the motivation for studying the dual convergence for a purely primal method as (Penalty-PPA) is two-folded. On the one hand, dual solutions are interesting on their own: they are useful for sensitivity analysis, and in some applications they have an actual modelling interpretation. On the other hand, such a dual convergence result provides a theoretical basis for the following stopping rule for (Penalty-PPA): at each iteration we find x^k, compute (2) and verify whether the pair (x^k, λ^k) satisfies a relaxed version of the Karush-Kuhn-Tucker conditions. If the answer is positive, then STOP. In this paper we illustrate some of these aspects through simple numerical illustrations.

The paper is organized as follows: in section §2, we describe some of the algorithms which can be set in the framework of (Penalty-PPA), and give the basic assumptions we make on the behaviour of the paramaters and the generated sequences $(x^k)_k$ and $(g^k)_k$. Then, in section §3, we show how to associate a dual sequence $(\lambda^k)_k$ to those generated by (Penalty-PPA), under some additional hypotheses. In section §4, we state and prove our main result, that is the convergence of the dual sequence $(\lambda^k)_k$ to the θ^*-center of the dual problem associated to (P). In particular, we give a simple proof and make some comments in the linear case. Finally, we give simple numerical illustrations of our results in section §5, where we also discuss briefly on an effective implementation of (Penalty-PPA).

2 Unified framework for penalty proximal point algorithms

The implicit iterative scheme given by (Penalty-PPA) can be viewed as a generic algorithm, which unifies several methods that have been considered in the literature. Indeed, in the specific case where we impose $\eta^k = e^k = 0$, then (Penalty-PPA) amounts to

$$x^k = x^{k-1} - h_k g^k, \quad g^k \in \partial_{\varepsilon_k} f_{r_k}(x^k). \tag{3}$$

This is exactly the method whose primal convergence is studied in [13] under some hypotheses, which imply in particular that $r_k \to 0$ together with the summability condition

$$\sum \varepsilon_k h_k < +\infty. \tag{4}$$

See also [8, 21, 22]. Notice that if in addition we require that $\varepsilon_k = 0$, then (3) corresponds to the equation

$$\nabla f_{r_k}(x^k) + \frac{1}{h_k}(x^k - x^{k-1}) = 0, \tag{5}$$

and, by convexity, this means that x^k must solve the auxiliary *unconstrained* problem

$$\min_{x \in \mathbb{R}^d} \left\{ f_{r_k}(x) + \frac{1}{2h_k}\|x - x^{k-1}\|^2 \right\}.$$

When $\varepsilon_k = 0$, it is natural to relax (5) by introducing an error criterion of the type

$$\|\nabla f_{r_k}(x^k) + \frac{1}{h_k}(x^k - x^{k-1})\| \leq \delta_k$$

for a given tolerance $\delta_k > 0$, or equivalently we seek x^k such that

$$x^k = x^{k-1} - h_k \nabla f_{r_k}(x^k) + \eta^k,$$

for any η^k satisfying

$$\|\eta^k\| \le \delta_k h_k. \tag{6}$$

Of course, the last algorithm is a specialization of (Penalty-PPA) with $\varepsilon_k = 0$ and $e^k = 0$. More generally, for $\varepsilon_k > 0$ the following scheme can be considered

$$x^k = x^{k-1} - h_k g^k + \eta^k, \quad g^k \in \partial_{\varepsilon_k} f_{r_k}(x^k)$$

where η^k is required to satisfy (6). See [5, 15] for primal convergence results under the hypothesis $\sum \delta_k h_k < +\infty$ so that $\sum \|\eta^k\| < +\infty$; in particular, $\eta^k \to 0$ as $k \to +\infty$.

In order to impose less stringent conditions for the errors, some variants of the proximal point algorithm have been proposed in the literature. For instance, following [23, 26], a two-steps algorithm can be can considered, where an inexact proximal iteration is first performed to find an auxiliary point z^k such that

$$z^k = x^{k-1} - h_k g^k + \xi^k, \quad g^k \in \partial_{\varepsilon_k} f_{r_k}(z^k), \tag{7}$$

for some error ξ^k, next the current iterate x^{k-1} is updated through

$$x^k = x^{k-1} - \beta_k g^k \text{ with } \beta_k = \langle g^k, x^{k-1} - z^k \rangle / \|g^k\|^2. \tag{8}$$

The latter is a *projection step*. Indeed, (8) can be written as $x^k = P_k x^{k-1}$, where $P_k \colon \mathbb{R}^d \to \mathbb{R}^d$ is the orthogonal projection onto the hyperplane $\{x \in \mathbb{R}^d \mid \langle g^k, x - z^k \rangle = 0\}$. Such a *hybrid* projection-proximal algorithm has the advantage to ensure primal convergence under a fixed relative error condition, namely, the error ξ^k is required to satisfy

$$\|\xi^k\| \le \sigma \sqrt{\|z^k - x^k\|^2 + h_k^2 \|g^k\|^2}, \tag{9}$$

where $\sigma \in [0, 1[$ is given and remains fixed for all k. See [4] for results in this direction in the case of penalty methods.

It is not difficult to show that (7)-(9) yield a specialization of (Penalty-PPA) for certain η^k and e^k, both converging to zero as $k \to \infty$. In fact, under fairly general conditions, as a by-product of the convergence analysis in [4], it follows that in this case $\sum h_k^2 \|g^k\|^2 < +\infty$ and $\sum \|z^k - x^k\|^2 < +\infty$, hence a posteriori we have $\sum \|\xi^k\|^2 < +\infty$ and moreover we can write z^k as

$$z^k = x^k + e^k$$

for some e^k with $\sum \|e^k\|^2 < +\infty$. By (7), we have that x^k satisfies (Penalty-PPA) for $\eta^k := \xi^k - e^k$. In a similar way, we note that (8) can be written directly as (Penalty-PPA) where β_k takes the place of h_k, $\eta_k := 0$ and $z^k = x^k + e^k$. In any case, for the purpose of this paper, it will be only important to know that $\eta_k \to 0$ and $e^k \to 0$. Similar considerations hold for a hybrid *extragradient*-proximal method [26] where the projection step (8) is replaced with $x^k = x^{k-1} - h_k g^k$, and the error ξ^k is required to satisfy a fixed relative error tolerance analogous to (9). The coupling of this variant with penalty methods have been recently investigated in [11].

We finally notice that the above discussion also applies to penalty methods with two parameters as developed in [19]. These methods are based on the same scheme as (Penalty-PPA), but including the penalty function θ appearing in f_{r_k} replaced by $\beta_k \theta$, where the varying parameter β_k being increased on each iteration for which the iterate x^k is not feasible. It is proved in [19] that in the convex case, and assuming that the Slater condition holds, the sequence $(\beta_k)_k$ generated by the algorithm is stationary. In this respect, our asymptotic analysis of the next sections does apply (with $(\lim \beta_k)\theta$ replacing θ) to this type of method.

Under appropriate conditions on the data and the algorithm parameters, the iterative schemes described by (Penalty-PPA) generate a primal sequence $(x^k)_k$ whose cluster points belong to the optimal set $S(P)$. From now on, we shall assume that the sequences $(x^k)_k$ and $(g^k)_k$ generated by (Penalty-PPA) satisfy

$$x^k \to x^\infty \text{ and } g^k \to 0 \text{ as } k \to +\infty \tag{H$_0$}$$

for some $x^\infty \in S(P)$. In view of all known primal convergence results (see, for instance, [4, 5, 13, 15, 19]), natural but not sufficient assumptions for (H$_0$) to hold are the following:

$$r_k,\, \varepsilon_k \to 0 \quad \text{as } k \to +\infty, \tag{H$_1$}$$

$$\eta^k,\, e^k \to 0 \quad \text{as } k \to +\infty, \tag{H$_2$}$$

as well as

$$\exists h > 0,\, \forall k \geq 0, \qquad h_k \geq h. \tag{H$_3$}$$

Remark 2.1. As part of the hypotheses stated in the introduction, we do assume throughout the paper that all the functions f_i are differentiable (in fact, in view of (H$_0$), local differentiability in the neighbourhood of x^∞ would be sufficient). With this respect, the assumption $g^k \to 0$ made in (H$_0$) is natural; for example, when the parametrization $(r_k)_k$ is slow, this amounts to follow the central trajectory $x(r)_r$. Moreover, this hypothesis is compatible with (H$_2$,H$_3$) in the sense that since $g^k \to 0$, we may assume that $(h_k)_k$ is bounded from below by a positive constant. This is no more valid in the non-differentiable case, as the following example shows.

Example 2.2. Consider the following problem, where the functions f_i are not differentiable at x^∞:

$$(P) \qquad \min\{|x| \,:\, x \in \mathbb{R}, |x| \leq 1\}.$$

Of course, the only solution of (P) is $x^\infty = 0$, but if for example one applies (Penalty-PPA) with $\theta = \exp$ and $\varepsilon_k \equiv 0$, then the assumption $g^k \to 0$ may only hold if the sequence $(g^k)_k$ is stationary with $x^k + e^k = 0$ at some iteration. We notice that in this example, it should be assumed that $h_k \to 0$, so that, $(x^k)_k$ may converge to x^∞.

Thus, in this paper, we take for granted that primal convergence holds, therefore we will focus on dual convergence.

3 Preliminaries on duality in penalty methods

3.1 Approximate multipliers

Let us introduce the dual problem associated with (P), which is

$$(D) \qquad \max_{\lambda \geq 0} \left\{ p(\lambda) := \inf_{x \in \mathbb{R}^d} L(x, \lambda) \right\}$$

where the Lagragian function $L\colon \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ is given by

$$L(x,\lambda) := f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x). \tag{10}$$

Since $S(P)$ is supposed to be nonempty and bounded, the dual optimal set $S(D)$ is also nonempty. The *Karush-Kuhn-Tucker conditions* associated with (P), consist on the following set of equations and inequalities:

$$(KKT) \qquad \begin{cases} \nabla_x L(x,\lambda) = 0, & \\ \lambda_i \geq 0,\ f_i(x) \leq 0, & \text{for all } i \in \{1,\ldots,m\}, \\ \sum_{i=1}^{m} \lambda_i f_i(x) = 0. & \end{cases}$$

It is well known that a pair $(\hat{x},\hat{\lambda}) \in \mathbb{R}^d \times \mathbb{R}^m$ satifies (KKT) if and only if $\hat{x}$ and $\hat{\lambda}$ are optimal solutions to (P) and (D), respectively.

Next, notice that the optimality condition for (P_r) can be written as

$$0 = \nabla f_r(x(r)) \;=\; \nabla f_0(x(r)) + \sum_{i=1}^{m} \lambda_i(r) \nabla f_i(x(r)) = \nabla_x L(x(r),\lambda(r))$$

where the multiplier vector $\lambda(r) \in \mathbb{R}^m_{++} := \{\lambda \in \mathbb{R}^m \mid \lambda_i > 0,\ i = 1,...,m\}$ is defined explicitly in terms of $x(r)$ by

$$\lambda_i(r) := \theta'\left(\frac{f_i(x(r))}{r}\right), \quad i = 1,...,m. \tag{11}$$

Notice that $\lambda(r)$ is the unique solution for the following problem:

$$(D_r) \qquad \max_{\lambda \geq 0} \left\{ p(\lambda) - r \sum_{i=1}^{m} \theta^*(\lambda_i) \right\}.$$

Here, $\theta^*(\lambda) := \sup_u \{\lambda u - \theta(u)\}$ is the *Fenchel conjugate* of θ, which plays the role of a barrier function for the positivity constraint. Indeed, for the examples mentioned in §1, we have: $\theta_1^*(\lambda) = \lambda \log \lambda - \lambda$ if $\lambda \geq 0$, $\theta_1^*(\lambda) = \infty$ otherwise; $\theta_2^*(\lambda) = -1 - \log \lambda$ and $\theta_3^*(\lambda) = 1/\sqrt{\lambda} - \sqrt{\lambda}$ if $\lambda > 0$, $\theta_2^*(\lambda) = \theta_3^*(\lambda) = \infty$ otherwise. Furthermore, the dual optimal path $\lambda(r)$ converges to some $\lambda^\theta \in S(D)$ as $r \to 0^+$. See, for instance, [6, 7, 15].

Similarly, we can associate with the primal sequence $(x^k)_k$ obtained by iterating (Penalty-PPA), the sequence $(\lambda^k)_k$ given by (2). Notice that for any k the vector λ^k is well defined because $g^k \in \partial_{\varepsilon_k} f_{r_k}(x^k + e^k)$, so that, $x^k + e^k$ is in the domain of f_{r_k}. The following result illustrates why $(\lambda^k)_k$ may be considered as a *multiplier sequence* associated with the primal sequence $(x^k)_k$.

Lemma 3.1. *Assume that* (H_0)-(H_3) *hold, as well as*

$$\frac{\varepsilon_k}{r_k} \to 0 \quad \text{as} \quad k \to +\infty \tag{H_4}$$

If $\bar{\lambda}$ is a cluster point of $(\lambda^k)_k$ as $k \to +\infty$, then the pair $(x^\infty, \bar{\lambda})$ satisfies (KKT).

Proof. When $\varepsilon_k > 0$, by the Brøndsted-Rockafellar Theorem [10], there exists $\tilde{x}^k$ such that

$$\|x^k + e^k - \tilde{x}^k\| \le \sqrt{\varepsilon_k\, r_k} \quad \text{and} \quad \|g^k - \zeta^k\| \le \sqrt{\frac{\varepsilon_k}{r_k}}, \tag{12}$$

where $\zeta^k = \nabla f_{r_k}(\tilde{x}^k)$. When $\varepsilon_k = 0$, it may be set $\tilde{x}^k := x^k + e^k$ and $\zeta^k := g^k$. For $i \in \{1, \ldots, m\}$ and $k \ge 0$, we set $\tilde{\lambda}_i^k := \theta'[f_i(\tilde{x}^k)/r_k]$ and notice that

$$\langle \nabla f_i(\tilde{x}^k)\,,\, \frac{x^k + e^k - \tilde{x}^k}{r_k} \rangle \;\le\; \frac{f_i(x^k + e^k) - f_i(\tilde{x}^k)}{r_k} \;\le\; \langle \nabla f_i(x^k + e^k)\,,\, \frac{x^k + e^k - \tilde{x}^k}{r_k} \rangle.$$

It then follows from (H_4), that the sequences $(\lambda^k)_k$ and $(\tilde{\lambda}^k)_k$ have the same asymptotic behaviour as $k \to +\infty$. In particular, $\bar{\lambda}$ is a cluster point of $\{\tilde{\lambda}^k : k \to +\infty\}$. By definition, for all i, k we have $\tilde{\lambda}_i^k > 0$, so that, $\bar{\lambda} \ge 0$. Moreover, it holds

$$\begin{aligned} \zeta^k \;=\; \nabla f_{r_k}(\tilde{x}^k) \;&=\; \nabla f_0(\tilde{x}^k) + \sum_{i=1}^{m} \theta'[f_i(\tilde{x}^k)/r_k]\, \nabla f_i(\tilde{x}^k) \\ &=\; \nabla f_0(\tilde{x}^k) + \sum_{i=1}^{m} \tilde{\lambda}_i^k\, \nabla f_i(\tilde{x}^k) \;=\; \nabla_x L(\tilde{x}^k, \tilde{\lambda}^k). \end{aligned}$$

From (H_0) and (H_2), we then conclude that $\tilde{x}^k \to x^\infty$ and $\zeta^k \to 0$, so that, in the limit we have that $0 = \nabla_x L(x^\infty, \bar{\lambda})$. It then remains to notice that

$$\bar{\lambda}_i \;=\; \lim_{k \to +\infty} \theta' \left[f_i(\tilde{x}^k)/r_k \right] \;=\; 0$$

for any index i such that $f_i(x^\infty) < 0$ because $r_k \to 0$ and $\theta'(u) \to 0$ as $u \to -\infty$. As a consequence, the complementarity holds:

$$\sum_{i=1}^{m} \bar{\lambda}_i f_i(x^\infty) = 0,$$

which concludes the proof. □

3.2 Boundedness and convergence to the dual optimal set

From now on, we assume that

$$S(D) \text{ is nonempty and bounded.} \tag{H_5}$$

A sufficient condition for the latter is the Slater condition (1), which we assume in the case where θ is such that $\kappa = 0$.

Proposition 3.2. *Under the assumptions* (H_0)-(H_5), *the dual sequence* $(\lambda^k)_k$ *is bounded and, even more,* $\operatorname{dist}(\lambda^k, S(D)) \to 0$ *as* $k \to +\infty$.

Proof. As in the proof of Lemma 3.1, when $\varepsilon_k > 0$, it is more natural to study the sequence $(\tilde{\lambda}^k)_k$ associated with $(\tilde{x}^k)_k$ obtained via the Brøndsted-Rockafellar Theorem, which has the same asymptotic behaviour as $(\lambda^k)_k$. If $\varepsilon_k = 0$, then we simply identify $(\tilde{\lambda}^k)_k$ with $(\lambda^k)_k$.

Next, recall from the proof of Lemma 3.1 that

$$\zeta^k \;=\; \nabla f_0(\tilde{x}^k) + \sum_{i=1}^m \tilde{\lambda}_i^k \nabla f_i(\tilde{x}^k)$$

with $\tilde{x}^k \to x^\infty$ and $\zeta^k \to 0$. If the sequence $(\tilde{\lambda}^k)_k$ were unbounded, then dividing by $\sum_{i=1}^m \tilde{\lambda}_i^k$, letting $k \to +\infty$ and taking a subsequence if necessary, we would deduce that there exist $\alpha_1, ..., \alpha_m \geq 0$ with $\sum_{i=1}^m \alpha_i = 1$ such that

$$0 \;=\; \sum_{i=1}^m \alpha_i \nabla f_i(x^\infty).$$

As a result of $\tilde{\lambda}_i^k \to 0$ for all i such that $f_i(x^\infty) < 0$, we would have $\alpha_i = 0$ for any of such i's. This implies that $S(D) = S(D) + \mathbb{R} \cdot \alpha$, for some $\alpha \neq 0 \in \mathbb{R}^m$, contradicting the boundedness of $S(D)$.

As a consequence, $(\tilde{\lambda}^k)_k$ is bounded, and it results from Lemma 3.1 that any of its cluster points is a Lagrange multiplier for x^∞, which implies that $\operatorname{dist}(\tilde{\lambda}^k, S(D)) \to 0$. □

3.3 The θ^*-center

In the next section, we will study sufficient conditions ensuring that $(\lambda^k)_k$ converges to a particular solution of (D), which, in fact will be the so-called *θ^*-center* of the dual optimal set $S(D)$. In this section, we introduce this particular dual solution and give some of its relevant properties for our asymptotic analysis.

We begin by setting

$$I \;:=\; \{i \mid \exists \mu \in S(D),\, \mu_i > 0\}\,.$$

Notice that since $S(D)$ is convex, there exists $\mu \in S(D)$ such that $\mu_i > 0$ for any $i \in I$.

Definition 3.3. *The θ^*-center of $S(D)$ is the unique $\lambda^\theta \in S(D)$ such that*

$$\sum_{i\in I} \theta^*(\lambda_i^\theta) \;=\; \min_{\mu \in S(D)} \left\{ \sum_{i\in I} \theta^*(\mu_i) \right\}.$$

The existence and uniqueness of the θ^*-center follow from (H_5) and the hypotheses made on θ. The θ^*-center was introduced and identified as the natural candidate for the asymptotic limit of the dual path associated with penalty methods in [6, 7, 15, 18].

The two following results are at the root of the convergence analysis of section §4.

Lemma 3.4. *Assume that $\lambda \in S(D)$ is such that*

$$\forall i \in I, \qquad \lambda_i \;=\; \theta'(\langle \nabla f_i(x^\infty), v\rangle)$$

for some $v \in \mathbb{R}^d$. Then $\lambda = \lambda^\theta$, i.e. *λ is the θ^*-center of $S(D)$.*

Proof. Since λ and λ^θ both belong to $S(D)$, and recalling that $\lambda_i = \lambda_i^\theta = 0$ for any $i \in \{1, \ldots, m\} \backslash I$, we get

$$\sum_{i\in I} \lambda_i \nabla f_i(x^\infty) \;=\; -\nabla f_0(x^\infty) \;=\; \sum_{i\in I} \lambda_i^\theta \nabla f_i(x^\infty).$$

Next, we compute

$$\begin{aligned}\sum_{i\in I}\theta^*(\lambda_i^\theta)-\theta^*(\lambda_i) &\geq \sum_{i\in I}\theta^{*\prime}(\lambda_i)(\lambda_i^\theta-\lambda_i) \;=\; \sum_{i\in I}\langle\nabla f_i(x^\infty),v\rangle(\lambda_i^\theta-\lambda_i)\\ &= \langle\sum_{i\in I}\lambda_i^\theta\nabla f_i(x^\infty)-\sum_{i\in I}\lambda_i\nabla f_i(x^\infty)\,,v\rangle \;=\; 0,\end{aligned}$$

where the first inequality is by convexity of θ^*, and then we have used $\theta^{*\prime}=\theta'^{-1}$. This implies that λ minimizes $\mu\mapsto\sum_{i\in I}\theta^*(\mu_i)$ over $S(D)$. By uniqueness, we deduce that $\lambda=\lambda^\theta$. ☐

Lemma 3.5. *Let us consider* $\Psi\colon\mathbb{R}^d\to\mathbb{R}\cup\{+\infty\}$ *defined by*

$$\Psi(v) \;:=\; \langle\nabla f_0(x^\infty),v\rangle+\sum_{i\in I}\theta(\langle\nabla f_i(x^\infty),v\rangle). \tag{13}$$

If $\mathbb{F}:=\mathit{Span}\{\nabla f_i(x^\infty)\mid i\in I\}$*, then the restriction of* Ψ *to* $\mathbb{F}$ *is coercive, strictly convex and has a unique minimizer* v^∞ *on* $\mathbb{F}$*. Moreover, we have*

$$\forall i\in I,\;\lambda_i^\theta \;=\; \theta(\langle\nabla f_i(x^\infty),v^\infty\rangle). \tag{14}$$

Proof. For $v\in\mathbb{F}$, the value of the recession function Ψ_∞ of Ψ at v is given by

$$\begin{aligned}\Psi_\infty(v) &= \langle\nabla f_0(x^\infty),v\rangle+\sum_{i\in I}\theta_\infty(\langle\nabla f_i(x^\infty),v\rangle)\\ &= \sum_{i\in I}\Big[\theta_\infty(\langle\nabla f_i(x^\infty),v\rangle)-\lambda_i^\theta\langle\nabla f_i(x^\infty),v\rangle\Big]\end{aligned}$$

where we used $\lambda^\theta\in S(D)$. The hypotheses made on θ yield that $\theta_\infty(s)=0$ whenever $s\leq 0$ and $\theta_\infty(s)=+\infty$, whenever $s>0$. Since $\lambda_i^\theta>0$ for any $i\in I$, it then follows that $\Psi_\infty(v)\in]0,+\infty]$ for any $v\in\mathbb{F}\setminus\{0\}$. The strict convexity of Ψ on $\mathbb{F}$ is inherited from that of θ, besides the existence and uniqueness of the minimizer v^∞ follows directly.

The optimality condition for the unique minimizer v^∞ of Ψ over $\mathbb{F}$ reads

$$-\nabla f_0(x^\infty) \;=\; \sum_{i\in I}\theta'(\langle\nabla f_i(x^\infty),v^\infty\rangle)\nabla f_i(x^\infty).$$

As a consequence, the vector $\lambda\in\mathbb{R}^m$ with coordinates $\lambda_i:=\theta'(\langle\nabla f_i(x^\infty),v^\infty\rangle)$ for $i\in I$ and $\lambda_i=0$ otherwise, belongs to $S(D)$. Therefore, (14) follows from Lemma 3.4. ☐

4 Convergence to the θ^*-center

4.1 The main convergence result

By virtue of the caracterization of the θ^*-center given in Lemma 3.4, as well as, the notions introduced in Lemma 3.5, we are now in position to state our main convergence result. From now on, $\mathrm{Proj}_{\mathbb{F}}(\cdot)$ denotes the usual projection on the vector subspace $\mathbb{F}$ defined in Lemma 3.5.

Theorem 4.1. *Let* (H$_0$)-(H$_5$) *hold, and assume that* $(\frac{r_k}{r_{k-1}})_{k\geq 1}$ *is a bounded sequence. Then, the dual sequence* $(\lambda^k)_k$ *converges to the* θ^**-center* λ^θ*, as* $k \to +\infty$*. Moreover, we also have*

$$\mathrm{Proj}_{\mathbb{F}}(w^k) \to v^\infty \quad as \quad k \to +\infty \tag{15}$$

where $w^k := \dfrac{(x^k + e^k) - x^\infty}{r_k}$*, and* v^∞ *is the unique optimal minimizer of* Ψ *over* $\mathbb{F}$*.*

The above convergence result generalizes that of [5, 15], which was devoted to the case where all the functions f_i were assumed to be affine (we refer to the following subsection §4.2 for more comments on this issue); we only require that the functions f_i be convex and (locally) differentiable around the limit point x^∞.

We notice that Theorem 4.1 in particular yields the following asymptotic behaviour for $(\lambda^k)_k$

$$\forall i \in I, \qquad \lambda_i^k = \theta'\left[\frac{f_i(x^\infty + r_k\, w^k)}{r_k}\right] \to \lambda_i^\theta = \theta'(\langle \nabla f_i(x^\infty), v^\infty \rangle),$$

as well as the asymptotic expansion

$$\mathrm{Proj}_{\mathbb{F}}(x^k + e^k) = \mathrm{Proj}_{\mathbb{F}}(x^\infty) + r_k v^\infty + o(r_k). \tag{16}$$

Remark 4.2. As noted in Remark 2.1, the differentiability assumption on the functions f_i is linked with the hypotheses (H$_0$) (for the part $g^k \to 0$) and (H$_3$), which appear to be necessary in our convergence analysis. On the other hand, our proof of Theorem 4.1 mainly relies on obtaining the asymptotic expansion (16). With this in mind, the differentiability assumption on the functions f_i is also quite natural for this first order expansion. As a consequence, relaxing this hypothesis may involve quite a different approach.

One of the main difficulties of the proof of Theorem 4.1, which we postpone to §4.3, is that the sequence $(w^k) = (\frac{(x^k+e^k)-x^\infty}{r_k})$ may be unbounded in $\mathbb{R}^d$ when strict complementarity does not hold. The following example shows this fact and also illustrates why the asymptotic expansion (16) only holds in $\mathbb{F}$.

Example 4.3. Consider the following simple example in dimension 2:

$$(P_{ex}) \qquad \min\left\{x_1 \,:\, x_1^2 + (x_2-1)^2 \leq 1, \quad -2\,x_1 \leq 2 \quad x_2 - x_1 \leq 2\right\}$$

that is

$$f_0(x) = x_1;\;\; f_1(x) = x_1^2 + (x_2-1)^2 - 1;\;\; f_2(x) = -2\,x_1 - 2 \;\text{ and }\; f_3(x) = x_2 - x_1 - 2.$$

Then, $S(P_{ex}) = \{x^\infty\}$ with $x^\infty := (-1,1)$, $S(D_{ex}) = \{(t, \frac{1}{2} - t, 0) : t \in [0, \frac{1}{2}]\}$ so that, $I = \{1,2\}$ and $\mathbb{F} = Span(\nabla f_1(x^\infty), \nabla f_2(x^\infty)) = \mathbb{R} \times \{0\}$. Take $\theta = \exp$ and let $(x^k)_k$ be generated by (Penalty-PPA). Under the hypotheses of Theorem 4.1, we infer that the associated dual path $(\lambda^k)_k$ converges to the θ^*-center λ^θ given by

$$\lambda^\theta := \mathrm{argmin}\,\{\theta^*(\lambda_1) + \theta^*(\lambda_2) \,:\, \lambda \in S(D_{ex})\} = \left(\frac{1}{4}, \frac{1}{4}, 0\right)$$

while $\mathrm{Proj}_{\mathbb{F}}(w^k)$ converges to $v^\infty := (s_\theta, 0)$, where

$$s_\theta \;:=\; \mathrm{argmin}\,\{s + 2\,\theta(-2\,s) \;:\; s \in \mathbb{R}\} = \ln(2).$$

This implies that

$$w_1^k \;=\; \frac{x_1^k + e_1^k - x_1^\infty}{r_k} \;=\; \frac{x_1^k + e_1^k + 1}{r_k} \;\to\; \ln(2) \quad \text{as} \quad k \to +\infty.$$

On the other hand, since the Lagrange multiplier for f_3 is always 0, we infer that $\lambda_3^k = \theta'[f_3(x^k + e^k)/r_k] \;\to\; 0$, that is

$$\frac{f_3(x^k + e^k)}{r_k} \;=\; w_2^k - w_1^k \;\to\; -\infty \quad as \quad k \to +\infty.$$

Since $(w_1^k)_k$ converges, we conclude that $w_2^k \to -\infty$. See §5.2 for a numerical illustration of that example.

The rest of the section is devoted to the proof of Theorem 4.1. For the reader's convenience, we shall first detail the proof in the linear case, where its technique appears more clearly. Finally, we shall turn to the proof in the general case.

4.2 The linear case

In this section, we consider the case where all the functions f_i are assumed to be affine, so that, (P) is the following standard linear programming problem

$$(P) \qquad \min_{x \in \mathbb{R}^d} \left\{ \langle a_0, x\rangle \;:\; \forall i \in \{1, \dots m\}, \langle a_i, x\rangle \le b_i \right\}.$$

We consider a sequence $(x^k)_k$ generated by (Penalty-PPA), then the associated dual path $(\lambda^k)_k$ is given by

$$\forall i \in \{1, \dots, m\}, \qquad \lambda_i^k = \theta' \left[(\langle a_i, x^k + e^k\rangle - b_i)/r_k \right].$$

We notice that for any $i \in I$, the complementarity condition yields $\langle a_i^t, x^\infty\rangle = b_i$, so that, we have

$$\forall i \in I, \qquad \lambda_i^k = \theta' \left[\langle a_i, w^k\rangle \right] \tag{17}$$

where $w^k := \dfrac{(x^k + e^k) - x^\infty}{r_k}$. Furthermore, the function Ψ defined in Lemma 3.5 is given by

$$\Psi(v) \;:=\; \langle a_0, v\rangle + \sum_{i \in I} \theta[\langle a_i, v\rangle],$$

and $\mathbb{F} = Span\{a_i \;:\; i \in I\}$. In this setting, Theorem 4.1 reads as follows.

Theorem 4.4. *Let* (H$_0$)-(H$_5$) *hold, and assume that* $(\frac{r_k}{r_{k-1}})_{k \ge 1}$ *is a bounded sequence. Then, the dual sequence* $(\lambda^k)_k$ *converges to the* θ^**-center* λ^θ*, as* $k \to +\infty$*. In particular, we have*

$$\forall i \in I, \qquad \langle a_i, w^k\rangle \;\to\; \langle a_i, v^\infty\rangle \tag{18}$$

where $w^k := \dfrac{(x^k + e^k) - x^\infty}{r_k}$*, and* v^∞ *is the unique optimal minimizer of* Ψ *over* $\mathbb{F}$*.*

Remark 4.5. Notice that, thanks to Lemma 3.5 and (17), the property (18) is in fact equivalent to the convergence of $(\lambda^k)_k$ to the θ^*-center λ^θ. The difficulty for the general nonlinear case lies mainly in the fact that the equivalence between the convergence of $(\lambda^k)_k$ to λ^θ and (15), does not hold *a priori* anymore.

Similar results to Theorem 4.4 about the convergence of the dual sequence $(\lambda^k)_k$ were obtained in [5] and [15], in the setting of linear programming. More precisely, in [5] a similar dual convergence result is proved for the particular case where $\theta \equiv \exp(\cdot)$ is the exponential penalty, and replacing (H_3) by the more general $\sum h_k = +\infty$, together with the following additional hypotheses: $(r_k)_k$ is non-increasing, $(\frac{r_{k-1}-r_k}{r_{k-1}h_k})$ is bounded, $\frac{\varepsilon_k}{h_k} \to 0$, and $e^k \equiv 0$ in (H_2). In [15], two dual convergence results are proved for a wide class of penalty barrier functions. On the one hand, assuming that the penalty θ is bounded from below, the convergence result is obtained under the same hypotheses on the parameters as in [5]. On the other hand, allowing θ to be unbounded from below, dual convergence is proved with the additional assumptions that $\varepsilon_k \equiv 0$ and $\eta^k \equiv e^k \equiv 0$ in (H_2). Notice that in Theorem 4.4, we assume that h_k is bounded away from zero, but we have proved the convergence including the parameters $\eta^k, e^k \in \mathbb{R}^d$, $\varepsilon_k \geq 0$, and for general penalty functions θ.

We shall now give a proof of Theorem 4.4, that we could in fact omit since this result is a straightforward corollary of Theorem 4.1 for the special setting of linear programming. Nevertheless, we believe that the fundamental arguments for the proof of Theorem 4.1 appear clearly in this simpler setting, therefore, we propose a short proof in this case. For the sake of simplification, we shall also restrict the proof to the case where $\varepsilon_k \equiv 0$. Let us mention as well, that our proof is much simpler than those in [5, 15].

Proof of Theorem 4.4 with $\varepsilon_k \equiv 0$. Thanks to Proposition 3.2, we get that $\lambda_i^k \to 0 = \lambda_i^\theta$ for any $i \notin I$. Thus, we shall study the convergence of the sequences $(\lambda_i^k)_k$ to λ_i^θ for $i \in I$. By Remark 4.5, it is then sufficient to prove that

$$w_{\mathbb{F}}^k \;:=\; Proj_{\mathbb{F}}(w^k) \;\to\; v^\infty. \tag{19}$$

We first claim that for all k it holds

$$\frac{1}{2}\|w_{\mathbb{F}}^k - v^\infty\|^2 - \frac{1}{2}\|w_{\mathbb{F}}^{k-1} - v^\infty\|^2 \;\leq\; \frac{h_k}{r_k}\left[\Psi(v^\infty) - \Psi(w_{\mathbb{F}}^k) + \langle \delta^k, w_{\mathbb{F}}^k - v^\infty\rangle\right] \tag{20}$$

with $\delta^k \to 0$ as $k \to +\infty$. In order to verify this, we first notice that

$$\begin{aligned} w^k - w^{k-1} &= \frac{1}{r_k}\left[(x^k - x^{k-1}) + (e^k - \frac{r_k}{r_{k-1}}e^{k-1}) + (1 - \frac{r_k}{r_{k-1}})(x^{k-1} - x^\infty)\right] \\ &= \frac{1}{r_k}\left[x^k - x^{k-1} + \delta_1^k\right] \end{aligned}$$

where $\delta_1^k \to 0$ as $k \to +\infty$. On the other hand, by definition of (Penalty-PPA), it follows that

$$\begin{aligned} x^k - x^{k-1} &= -h_k \nabla f_{r_k}(x^k + e^k) + \eta^k \\ &= -h_k \left(c + \sum_{i\in I} \theta'[\langle a_i, w^k\rangle] a_i + \sum_{i\notin I} \lambda_i^k a_i - \eta_k/h_k\right) \\ &= -h_k \left(\nabla\Psi(w_{\mathbb{F}}^k) + \sum_{i\notin I} \lambda_i^k a_i - \eta_k/h_k\right) \;=\; -h_k\left(\nabla\Psi(w_{\mathbb{F}}^k) + \delta_2^k\right) \end{aligned}$$

where $\delta_2^k \to 0$ as $k \to +\infty$ and we used $\langle a_i, w^k\rangle = \langle a_i, w_{\mathbb{F}}^k\rangle$ for all $i \in I$. Thus, we obtain

$$w^k - w^{k-1} \;=\; -\frac{h_k}{r_k}\left[\nabla\Psi(w_{\mathbb{F}}^k) + \frac{\delta_1^k}{h_k} + \delta_2^k\right] \;=\; -\frac{h_k}{r_k}\left[\nabla\Psi(w_{\mathbb{F}}^k) + \delta^k\right] \tag{21}$$

where $\delta^k \to 0$ as $k \to +\infty$. Next, we may compute

$$\begin{aligned}\frac{1}{2}\|w_{\mathbb{F}}^k - v^\infty\|^2 - \frac{1}{2}\|w_{\mathbb{F}}^{k-1} - v^\infty\|^2 \;&=\; \langle w_{\mathbb{F}}^k - w_{\mathbb{F}}^{k-1}, \frac{w_{\mathbb{F}}^k + w_{\mathbb{F}}^{k-1}}{2} - v^\infty\rangle \\ &=\; \langle w_{\mathbb{F}}^k - w_{\mathbb{F}}^{k-1}, w_{\mathbb{F}}^k - v^\infty\rangle - \frac{1}{2}\|w_{\mathbb{F}}^k - w_{\mathbb{F}}^{k-1}\|^2 \\ &=\; \frac{h_k}{r_k}\langle \nabla\Psi(w_{\mathbb{F}}^k) + \delta^k, v^\infty - w_{\mathbb{F}}^k\rangle - \frac{1}{2}\|w_{\mathbb{F}}^k - w_{\mathbb{F}}^{k-1}\|^2,\end{aligned}$$

from which (20) follows by the convexity of Ψ.

From the coercivity of Ψ over $\mathbb{F}$ (see Lemma 3.5), we also get that

$$\forall\rho > 0,\ \exists\gamma > 0,\ \forall w \in \mathbb{F}, \qquad \|w - v^\infty\| \geq \rho \Rightarrow \Psi(w) - \Psi(v^\infty) \geq \gamma.$$

Since $\delta^k \to 0$, we deduce from (20) that

$$\left\{\begin{array}{l}\text{for any } \rho > 0 \text{ there exists } \Gamma > 0 \text{ such that for } k \text{ large enough:} \\ \|w_{\mathbb{F}}^k - v^\infty\| \geq \rho \quad \Rightarrow \quad \frac{1}{2}\|w_{\mathbb{F}}^k - v^\infty\|^2 - \frac{1}{2}\|w_{\mathbb{F}}^{k-1} - v^\infty\|^2 \;\leq\; -\frac{h}{r_k}\Gamma\end{array}\right. \tag{22}$$

Since $r_k \to 0$, a classical argument yields $\|w_{\mathbb{F}}^k - v^\infty\| \to 0$, which proves the result.

□

4.3 Proof in the general case

We divide the proof in a series of steps, where Steps 1 and 4 are the analogous of (20) and (22) in this general setting.

Step 0. Proposition 3.2 implies that $\lambda_i^k \to 0$ for any $i \notin I$, so that, we only need to study the convergence of those coordinates of $(\lambda^k)_k$ with indices in I.

Following the proof of Lemma 3.1, we shall restrict our attention to the study of the sequences $(\tilde\lambda^k)_k$ and $(\tilde w^k)_k$, where

$$\tilde w^k \;:=\; \frac{\tilde x^k - x^\infty}{r_k} \quad \textit{and} \quad \tilde\lambda_i^k \;:=\; \theta'\left(\frac{f_i(\tilde x^k)}{r_k}\right) \;=\; \theta'\left(\frac{f_i(x^\infty + r_k\,\tilde w^k) - f_i(x^\infty)}{r_k}\right)$$

in which we recall that $f_i(x^\infty) = 0$ for any $i \in I$. It follows from the convexity of the functions f_i that

$$\forall i,\ \forall k,\ \forall y \in \mathbb{R}^d, \qquad \langle\nabla f_i(x^\infty), y\rangle \;\leq\; \frac{f_i(x^\infty + r_k\,y) - f_i(x^\infty)}{r_k} \;\leq\; \langle\nabla f_i(x^\infty + r_k\,y), y\rangle. \tag{23}$$

In particular, we have

$$\forall i \in \{0\} \cup I,\ \forall k, \qquad \langle\nabla f_i(x^\infty), \tilde w_{\mathbb{F}}^k\rangle \;\leq\; \frac{f_i(x^\infty + r_k\,\tilde w^k) - f_i(x^\infty)}{r_k} \;\leq\; \langle\nabla f_i(\tilde x^k), \tilde w_{\mathbb{F}^k}^k\rangle \tag{24}$$

where the sequences $(\tilde{w}^k_{\mathbb{F}})_k$ and $(\tilde{w}^k_{\mathbb{F}^k})_k$ are given by

$$\tilde{w}^k_{\mathbb{F}} := \mathrm{Proj}_{\mathbb{F}}(\tilde{w}^k) \quad \text{and} \quad \tilde{w}^k_{\mathbb{F}^k} := \mathrm{Proj}_{\mathbb{F}^k}(\tilde{w}^k)$$

with $\mathbb{F}^k := \mathit{Span}\{\nabla f_i(\tilde{x}^k) \mid i \in I \cup \{0\}\}$. Thanks to Lemma 3.5 and (24), it is sufficient to prove that $(\tilde{w}^k_{\mathbb{F}})_k$ and $(\tilde{w}^k_{\mathbb{F}^k})_k$ both converge to v^∞.

In our study of the convergence of $(\tilde{w}^k_{\mathbb{F}^k})_k$, we shall introduce the sequence $(v^\infty_{\mathbb{F}^k})_k$ given by

$$\forall k, \qquad v^\infty_{\mathbb{F}^k} := \mathrm{Proj}_{\mathbb{F}^k}(v^\infty)$$

and the family of functions $\Psi^k \colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ defined as

$$\Psi^k(v) = \frac{f_0(x^\infty + r_k\, v) - f_0(x^\infty)}{r_k} + \sum_{i\in I} \theta\left(\frac{f_i(x^\infty + r_k\, v) - f_i(x^\infty)}{r_k}\right).$$

Notice that it follows from (23) that

$$\forall (y^k)_k \in (\mathbb{R}^d)^{\mathbb{N}}, \qquad \left[y^k \to y\right] \;\Rightarrow\; \left[\Psi^k(y^k) \to \Psi(y)\right]. \tag{25}$$

Step 1. We claim that for all k it holds

$$\frac{1}{2}\|\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\|^2 - \frac{1}{2}\|\tilde{w}^{k-1}_{\mathbb{F}^{k-1}} - v^\infty_{\mathbb{F}^{k-1}}\|^2 \;\le\; \frac{h_k}{r_k}\left[\Psi^k(v^\infty_{\mathbb{F}^k}) - \Psi(\tilde{w}^k_{\mathbb{F}}) + \langle \delta^k, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle\right] \tag{26}$$

with $\delta^k \to 0$ as $k \to +\infty$.

We first compute

$$\begin{aligned}
&\frac{1}{2}\|\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\|^2 - \frac{1}{2}\|\tilde{w}^{k-1}_{\mathbb{F}^{k-1}} - v^\infty_{\mathbb{F}^{k-1}}\|^2 \\
= \;& \frac{1}{2}\langle \tilde{w}^k_{\mathbb{F}^k} - \tilde{w}^{k-1}_{\mathbb{F}^{k-1}} + v^\infty_{\mathbb{F}^{k-1}} - v^\infty_{\mathbb{F}^k}, \tilde{w}^k_{\mathbb{F}^k} + \tilde{w}^{k-1}_{\mathbb{F}^{k-1}} - (v^\infty_{\mathbb{F}^k} + v^\infty_{\mathbb{F}^{k-1}})\rangle \\
= \;& \frac{1}{2}\langle \tilde{w}^k_{\mathbb{F}^k} - \tilde{w}^{k-1}_{\mathbb{F}^{k-1}} + v^\infty_{\mathbb{F}^{k-1}} - v^\infty_{\mathbb{F}^k}, 2(\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k})\rangle - \frac{1}{2}\|(\tilde{w}^{k-1}_{\mathbb{F}^{k-1}} - \tilde{w}^k_{\mathbb{F}^k}) - (v^\infty_{\mathbb{F}^{k-1}} - v^\infty_{\mathbb{F}^k})\|^2 \\
\le \;& \langle \tilde{w}^k_{\mathbb{F}^k} - \tilde{w}^{k-1}_{\mathbb{F}^{k-1}}, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle + \langle v^\infty_{\mathbb{F}^{k-1}} - v^\infty_{\mathbb{F}^k}, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle \\
= \;& \langle \tilde{w}^k_{\mathbb{F}^k} - \tilde{w}^{k-1}_{\mathbb{F}^{k-1}}, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle + \frac{h_k}{r_k}\langle \delta^k_1, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle
\end{aligned}$$

where $\delta^k_1 := \frac{r_k}{h_k}(v^\infty_{\mathbb{F}^{k-1}} - v^\infty_{\mathbb{F}^k}) \to 0$ as $k \to +\infty$; indeed, by (H_3) the sequence $(h_k)_k$ is bounded from below by some positive h.

The first term of the right hand side is handled in the following way

$$\begin{aligned}
\langle \tilde{w}^k_{\mathbb{F}^k} - \tilde{w}^{k-1}_{\mathbb{F}^{k-1}}, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle &= \langle \tilde{w}^k_{\mathbb{F}^k} - \tilde{w}^{k-1}, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle + \langle \tilde{w}^{k-1} - \tilde{w}^{k-1}_{\mathbb{F}^{k-1}}, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle \\
&= \langle \tilde{w}^k - \tilde{w}^{k-1}, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle + \frac{h_k}{r_k}\langle \delta^k_2, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle
\end{aligned}$$

where we used that $\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k} \in \mathbb{F}^k$, and where $\delta^k_2 := \frac{1}{h_k}\frac{r_k}{r_{k-1}}(r_{k-1}\tilde{w}^{k-1} - r_{k-1}\tilde{w}^{k-1}_{\mathbb{F}^{k-1}}) \to 0$, as $k \to +\infty$.

We now compute

$$\begin{aligned}\tilde{w}^k - \tilde{w}^{k-1} &= \frac{1}{r_k}(\tilde{x}^k - \tilde{x}^{k-1}) + (\frac{1}{r_k} - \frac{1}{r_{k-1}})(\tilde{x}^{k-1} - x^\infty) \\ &= \frac{1}{r_k}(\tilde{x}^k - \tilde{x}^{k-1}) + \frac{h_k}{r_k}\delta_3^k\end{aligned}$$

where $\delta_3^k := \frac{1}{h_k}(1 - \frac{r_k}{r_{k-1}})(\tilde{x}^{k-1} - x^\infty) \to 0$, as $k \to +\infty$.
On the other hand, it follows by (12) and the definition of (Penalty-PPA) that

$$\begin{aligned}\tilde{x}^k - \tilde{x}^{k-1} &= x^k - x^{k-1} + e^k - e^{k-1} + \delta_4^k \;=\; -h_k g^k + h_k \eta^k + \delta_5^k \\ &= -h_k(\zeta^k + \delta_6^k) \;=\; -h_k\left(\nabla f_0(\tilde{x}^k) + \sum_{i=1}^m \tilde{\lambda}_i^k \nabla f_i(\tilde{x}^k) + \delta_6^k\right) \\ &= -h_k\left(\nabla f_0(\tilde{x}^k) + \sum_{i\in I} \tilde{\lambda}_i^k \nabla f_i(\tilde{x}^k) + \delta_7^k\right) \\ &= -h_k(\nabla \Psi^k(\tilde{w}^k) + \delta_7^k)\end{aligned}$$

where $\delta_j^k \to 0$ as $k \to +\infty$ for $j \in \{4, \dots, 7\}$.

The preceding yields

$$\frac{1}{2}\|\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\|^2 - \frac{1}{2}\|\tilde{w}^{k-1}_{\mathbb{F}^{k-1}} - v^\infty_{\mathbb{F}^{k-1}}\|^2 \;\leq\; \frac{h_k}{r_k}\left[\langle -\nabla\Psi^k(\tilde{w}^k), \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle + \langle \delta^k, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle\right]$$

with $\delta^k \to 0$ as $k \to +\infty$. Since $\nabla\Psi^k(\tilde{w}^k) \in \mathbb{F}^k$ we have

$$\langle -\nabla\Psi^k(\tilde{w}^k), \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle \;=\; \langle \nabla\Psi^k(\tilde{w}^k), v^\infty_{\mathbb{F}^k} - \tilde{w}^k\rangle \;\leq\; \Psi^k(v^\infty_{\mathbb{F}^k}) - \Psi^k(\tilde{w}^k).$$

It now remains to deduce from (23) that

$$-\Psi^k(\tilde{w}^k) \;\leq\; -\Psi(\tilde{w}^k) \;=\; -\Psi(\tilde{w}^k_{\mathbb{F}})$$

from which (26) follows.

Step 2. We now claim that for some $a > 0$, it holds

$$\forall k, \qquad \|\tilde{w}^k_{\mathbb{F}^k}\| \leq a\,(\|\tilde{w}^k_{\mathbb{F}}\| + 1) \quad \text{and} \quad \|\tilde{w}^k_{\mathbb{F}}\| \leq a\,(\|\tilde{w}^k_{\mathbb{F}^k}\| + 1). \tag{27}$$

Since the same arguments apply, we shall only prove that

$$\forall k, \qquad \|\tilde{w}^k_{\mathbb{F}^k}\| \leq a\,(\|\tilde{w}^k_{\mathbb{F}}\| + 1). \tag{28}$$

for some positive a. We make a proof by contradiction, and up to a change of index we assume that

$$\|\tilde{w}^k_{\mathbb{F}^k}\| \;\to\; +\infty \quad \text{and} \quad \frac{\|\tilde{w}^k_{\mathbb{F}}\|}{\|\tilde{w}^k_{\mathbb{F}^k}\|} \;\to\; 0 \quad \text{as } k \to +\infty.$$

Now (23) applied with $y = \tilde{w}^k$ yields that

$$\forall i \in I \cup \{0\},\; \forall k, \qquad \langle \nabla f_i(x^\infty), \tilde{w}^k_{\mathbb{F}}\rangle \;\leq\; \langle \nabla f_i(x^\infty + r_k\,\tilde{w}^k), \tilde{w}^k\rangle \;=\; \langle \nabla f_i(\tilde{x}^k), \tilde{w}^k_{\mathbb{F}^k}\rangle. \tag{29}$$

Thus, dividing by $\|\tilde{w}^k_{\mathbb{F}^k}\|$ we get

$$\forall i \in I \cup \{0\},\ \forall k, \qquad \langle \nabla f_i(x^\infty), \frac{\tilde{w}^k_{\mathbb{F}}}{\|\tilde{w}^k_{\mathbb{F}^k}\|} \rangle \ \le\ \langle \nabla f_i(\tilde{x}^k), \frac{\tilde{w}^k_{\mathbb{F}^k}}{\|\tilde{w}^k_{\mathbb{F}^k}\|} \rangle.$$

If we consider a cluster point $\tilde{v}$ of $(\frac{\tilde{w}^k_{\mathbb{F}^k}}{\|\tilde{w}^k_{\mathbb{F}^k}\|})_k$, then $\tilde{v}$ has norm 1 and belongs to $\mathbb{F}$, moreover

$$\forall i \in I \cup \{0\}, \qquad 0 \ \le\ \langle \nabla f_i(\tilde{x}^\infty), \tilde{v} \rangle.$$

Since it holds

$$\langle \nabla f_0(\tilde{x}^\infty), \tilde{v} \rangle \ =\ -\sum_{i \in I} \lambda_i^\theta \langle \nabla f_i(\tilde{x}^\infty), \tilde{v} \rangle \tag{30}$$

with $\lambda_i^\theta > 0$ for any $i \in I$, hence, we obtain

$$\forall i \in I \cup \{0\}, \qquad \langle \nabla f_i(\tilde{x}^\infty), \tilde{v} \rangle \ =\ 0$$

and then $\tilde{v} = 0$ since $\tilde{v} \in \mathbb{F}$, but this contradicts $\|\tilde{v}\| = 1$ and concludes this step.

Step 3. We infer from the preceding step the following estimate

$$\forall b > 0,\ \exists k_b \in \mathbb{N},\ \forall k \ge k_b, \qquad \|\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\| \ge b \Rightarrow \|\tilde{w}^k_{\mathbb{F}} - v^\infty\| \ge \frac{b}{2}. \tag{31}$$

We prove this by contradiction: assume that (31) fails for some $b > 0$, that is there exists an increasing function $\phi : \mathbb{N} \to \mathbb{N}$ such that

$$\forall k \in \mathbb{N}, \qquad \|\tilde{w}^{\phi(k)}_{\mathbb{F}^{\phi(k)}} - v^\infty_{\mathbb{F}^{\phi(k)}}\| \ge b \quad \textit{and} \quad \|\tilde{w}^{\phi(k)}_{\mathbb{F}} - v^\infty\| \le \frac{b}{2}.$$

It first follows from (27) that the sequence $(w^{\phi(k)}_{\mathbb{F}^{\phi(k)}})_k$ is bounded. We may then assume that $(\tilde{w}^{\phi(k)}_{\mathbb{F}^{\phi(k)}})_k$ converges to some limit $\tilde{v} \in \mathbb{F}$ while $(\tilde{w}^{\phi(k)}_{\mathbb{F}})_k$ converges to some limit $v \in \mathbb{F}$. Consequently, applying (29) with parameter $\phi(k)$ instead of k and then passing to the limit as $k \to +\infty$, gives

$$\forall i \in I \cup \{0\}, \qquad \langle \nabla f_i(x^\infty), v \rangle \ \le\ \langle \nabla f_i(x^\infty), \tilde{v} \rangle.$$

We conclude from these inequalities and (30) that $\tilde{v} = v$, since $\mathbb{F}$ is completely determined by $(\nabla f_i(x^\infty))_{i \in I \cup \{0\}}$. As $(\tilde{w}^{\phi(k)}_{\mathbb{F}^{\phi(k)}})_k$ and $(\tilde{w}^{\phi(k)}_{\mathbb{F}})_k$ have the same limit, yields the desired contradiction.

Step 4. We now conclude the proof of Theorem 4.1. To this end, we prove the following

$$\begin{cases} \text{for all } \rho > 0, \text{ there exists } \Gamma > 0 \text{ such that for } k \text{ large enough:} \\ \|\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\| \ge \rho \ \Rightarrow\ \frac{1}{2}\|\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\|^2 - \frac{1}{2}\|\tilde{w}^{k-1}_{\mathbb{F}^{k-1}} - v^\infty_{\mathbb{F}^{k-1}}\|^2 \ \le\ -\frac{\Gamma}{r_k} \end{cases} \tag{32}$$

Indeed, It directly follows from (32) that $\|\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\| \to 0$ as k goes to $+\infty$, hence $\tilde{w}^k_{\mathbb{F}^k} \to v^\infty$. Arguing as in Step 3, we infer that $\tilde{w}^k_{\mathbb{F}} \to v^\infty$ also holds, which ends the proof of Theorem 4.1.

In order to obtain (32), we fix $b > 0$ and notice the coercivity of Ψ over $\mathbb{F}$ obtained in Lemma 3.5, which yields that

$$\Gamma' \ :=\ \frac{1}{3} \inf\left\{ \frac{\Psi(v) - \Psi(v^\infty)}{\|v - v^\infty\|} \mid v \in \mathbb{F},\ \|v - v^\infty\| \ge \frac{\rho}{2} \right\} \ >\ 0.$$

For k large enough, if we assume that $\|\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\| \geq \rho$, then it follows from (26), (31) and the preceding that

$$\begin{aligned} & \frac{1}{2}\|\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\|^2 - \frac{1}{2}\|\tilde{w}^{k-1}_{\mathbb{F}^{k-1}} - v^\infty_{\mathbb{F}^{k-1}}\|^2 \\ \leq\ & \frac{h_k}{r_k}\left[\Psi^k(v^\infty_{\mathbb{F}^k}) - \Psi(v^\infty) + \Psi(v^\infty) - \Psi(\tilde{w}^k_{\mathbb{F}}) + \langle \delta^k, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle\right] \\ \leq\ & \frac{h_k}{r_k}\left[\Psi^k(v^\infty_{\mathbb{F}^k}) - \Psi(v^\infty) - 3\,\Gamma'\,\|\tilde{w}^k_{\mathbb{F}} - v^\infty_{\mathbb{F}^k}\| + \langle \delta^k, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle\right] \end{aligned}$$

Then, we infer from (25), (31) and the definition of δ_k that for k large enough, it holds

$$\Psi^k(v^\infty_{\mathbb{F}^k}) - \Psi(v^\infty) \leq \frac{\rho}{2}\Gamma' \quad \textit{and} \quad \langle \delta^k, \tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\rangle \ \leq\ \Gamma'\,\|\tilde{w}^k_{\mathbb{F}} - v^\infty_{\mathbb{F}^k}\|.$$

As a consequence, for k large enough we get

$$\|\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\| \geq \rho \quad \Rightarrow \quad \frac{1}{2}\|\tilde{w}^k_{\mathbb{F}^k} - v^\infty_{\mathbb{F}^k}\|^2 - \frac{1}{2}\|\tilde{w}^{k-1}_{\mathbb{F}^{k-1}} - v^\infty_{\mathbb{F}^{k-1}}\|^2 \ \leq\ -\frac{h_k}{r_k}\frac{\rho}{2}\Gamma'$$

from which (32) follows thanks to (H_3).

5 Numerical illustrations

In this section, we present some numerical illustrations of the theoretical results shown above. To this end, we first introduce the algorithm we used to implement the iterative scheme Penalty-PPA, and secondly, illustrate the convergence of the associated multipliers considering some examples.

5.1 The algorithm

For the numerical illustration, we have implemented the *hybrid projection proximal algorithm* (7)-(9) described in §(2). The main iteration of the algorithm is the following:

1. given x^{k-1} and r_k, solve (7),
2. obtain x^k via (8) and update r_k to r_{k+1}.

In order to solve (7), we note that when $\xi^k = 0$, the formula turns to the classical proximal point algorithm. Therefore, solving the equation (7) is equivalent to solve the system

$$\left\{ \begin{array}{lll} \nabla_z L(z,\lambda) + (z - x^{k-1})/h_k & = 0 & \\ \lambda_i - \theta'(f_i(z)/r_k) & = 0 & \forall i \in I(z,\lambda) \\ r_k \theta'^{-1}(\lambda_i) - f_i(z) & = 0 & \forall i \in \{1,\ldots,m\} \setminus I(z,\lambda) \end{array} \right. \tag{33}$$

This system is approximately solved using Newton's type iterations; being these *inner* iterations in (7) stopped when the relative error condition (9) is satisfied. Notice that the constraint $\lambda_i - \theta'(f_i(z)/r_k) = 0$ (for any $i \in \{1,\ldots,m\}$) is written in two equivalent forms in the above system. This way of splitting the constraints correspond to a well known technique of active set identification, and is introduced here to avoid the ill-conditioning of the system (for example, see [16, 17]). For the numerical illustrations presented hereafter, we have obtained good results with the following rule, in part inspired by the discussion in §3.2 of [17]: $i \in I(z, \lambda)$ if and only if $f_i(z)/r_k \leq \theta'^{-1}(r_k)$ and $\lambda_i < r_k$.

The penalty parameter r_k is updated using the feedback rule

$$r_{k+1} = \max\{\frac{1}{4} r_k, \min\{\|g^k\|^{(1.25)}, r_k\}\}. \quad (34)$$

This corresponds to a measure of how close is our iterate from the stationary point of f_{r_k}. We note that we are imposing bounds for updating the parameter in order to ensure that the sequence is decreasing, and also to avoid numerical instabilities when the parameter decreases too fast. We propose the factor 1.25 in the above equation in order to help the penalty parameter r_k be smaller than the norm of the gradient g^k: intuitively, this helps to strengthen the penalized part without making too many Newton's steps in (7). Without a special mention, the computations presented here use the rule (34), referred to as "feedback rule". As the results in Table 2 below show, this rule gives good results compared to fixed evolutions of $(r_k)_k$.

Finally, the main iteration stops when, for a given tolerance ε, we have

$$\max\left\{\frac{\|\nabla_x L(x^k, \lambda^k)\|}{\|\nabla f_0(x^k)\|}, \max_i\{f_i(x^k)\}, \max_i\{\left|\lambda_i^k f_i(x^k)\right|\}\right\} \leq \varepsilon, \quad (35)$$

which corresponds to a relaxed version of KKT conditions for the mathematical programming problem. In the criterion above, the division by $\|\nabla f_0(x^k)\|$ is intended to normalize the equation

$$\nabla_x L(x^k, \lambda^k) = 0 \quad \Leftrightarrow \quad \nabla f_0(x^k) = -\sum_{i=1}^{m} \lambda_i^k \nabla f_i(x^k).$$

Notice that the criterion (35) is validated by Theorem 4.1 which, in particular, yields that this maximum goes to 0 as $k \to \infty$.

For the numerical illustrations presented here, we choose $x^0 = 0 \in \mathbb{R}^n$, $\sigma = 0.95$, $h_k = 10$ for all $k \in \mathbb{N}$, $\theta = \exp(\cdot)$ as the penalty function, and $\varepsilon = 10^{-6}$ in (35) as the stopping criterion.

5.2 Examples

Our first illustration refers to Example 4.3 presented in §4. The results are shown in Figure 1, giving an emphasis upon the convergence of the dual sequence $(\lambda^k)_k$ to λ^θ, and on the distinct behaviours of the coordinates w_1^k and w_2^k. Notice that there is no explicit control on w^k in the algorithm, and, particularly, that the "divergent part" w_2^k may induce numerical instabilities in larger problems. Anyway, on this toy example in dimension 2, the lack of strict complementarity is well handled by our algorithm.

The second series of examples corresponds to the classical problem of minimizing the compliance of truss structures (for example see [2]). This problem consists in finding the distribution of bar's areas that minimizes the total amount of energy (compliance) of a truss structure carrying an external load. Mathematically, this problem can be stated as the following quadratic minimization problem:

$$\min_{x \in \mathbb{R}^d} \left\{\langle c, x\rangle \ : \ \frac{1}{2}\langle x, A_i x\rangle - \langle b, x\rangle \leq 0, \ i = 1, \ldots, m\right\}. \quad (36)$$

where $b, c \in \mathbb{R}^d$, and A_i, is a symmetric positive semidefinite matrix in $\mathbb{R}^d$, for $i = 1, \ldots, m$. It can be proved that the associated dual variables of this problem correspond to the bar's area of the

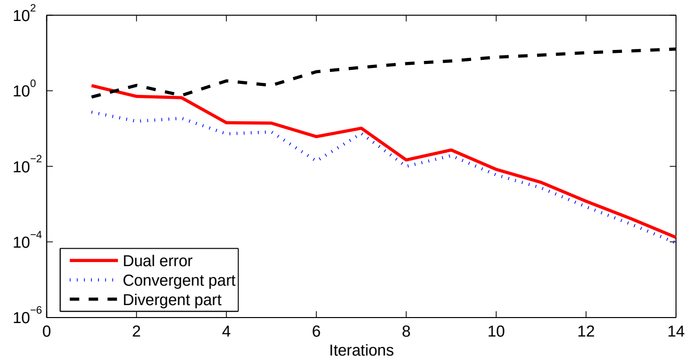

Figure 1: The continuous line (—) denotes the evolution of the dual residues ($\|\lambda^k - \lambda^\theta\| / \|\lambda^\theta\|$), the dotted line ($\cdots$) denotes the convergent direction ($\left|w_1^k - \ln 2\right| / \ln 2$), finally, the dashed line (– –) denotes the divergent part ($\left|w_2^k\right|$). At the stopping time, the value of r_k is $7 \cdot 10^{-8}$.

truss (see e.g. [1, 2]). Thus, any minimization algorithm implemented to solve (36) should ensure the convergence of the multipliers.

We illustrate the convergence of the algorithm (7)-(9) on various samples of this type of structural optimization problems. More precisely, we consider the following test problems:

- A cantilever type structure in a 3D space, see [24] Fig. 5.7, denoted here by **C**3D.
- A dome composed by four floors, see Example 2 Fig. 3 in [3], denoted here by **D**ome.
- A cantilever arm structure in a 2D space, similar to [2] §7.1, denoted here by **C**2D.
- A Michel type structure, similar to Fig. 1.1 in [20], denoted here by **MLP**.
- A cantilever type structure in a 3D space with a cross sectional area, similar to [9] Fig. 4.9 (b), denoted here by **C**3Db.

The results obtained are presented in Table 1. Notice that the stopping criterion (35) imposes that the couple (x^k, λ^k) returned by the algorithm satisfies the KKT system, which in particular validates the convergence of the dual variable λ^k to a dual solution.

problem	# variables	# constraints	iterations	# function evaluations
C3D	37	102	68	410
Dome	76	104	69	504
C2D	49	193	150	856
MLP	101	934	196	921
C3Db	145	912	111	1123

Table 1: Examples of small scale problems in structural optimization.

Following, we illustrate the convergence of the multipliers given by the algorithm (7)-(9) to the θ^*-center of the dual space, when the dual optimal set $S(D)$ is not a singleton. To do so, we consider the following reformulation of the problem **C**3D (presented in Table 1):

$$\min_{x\in\mathbb{R}^d}\left\{\langle c,x\rangle\,;\ \frac{1}{2}\left(\frac{1}{2}\langle x,B_ix\rangle-\langle b,x\rangle\right)\leq 0,\quad i=1,\ldots,2m\right\},\tag{37}$$

with $B_i := B_{i+m} := A_i$ for $i\in\{1,\ldots,m\}$. By duplicating each constraint, the dual problem of (37) admits multiple solutions, and its θ^*-center is $\bar\lambda^\theta=[\lambda^\theta,\lambda^\theta]$, where λ^θ is the θ^*-center of the original problem **C**3D. Then, we run the algorithm (7)-(9) by using different rules to update the penalty parameter r_k. The dual residues, the number of iterations and the total number of functional evaluations in the inner iterations (33) are reported in Table 2. It appears that the feedback rule (34) gives fairly good results compared to fixed parametrization rules for $(r_k)_k$.

Penalty parameter r_k	$\Vert\lambda^k-\bar\lambda^\theta\Vert/\Vert\bar\lambda^\theta\Vert$	# Iterations	# function evaluations
$1/k^2$	$2\cdot10^{-3}$	481	870
$1/k^3$	$5\cdot10^{-4}$	97	1108
$(0.8)^k$	$4\cdot10^{-5}$	72	507
feedback rule	$8\cdot10^{-5}$	83	502

Table 2: Behaviour of dual residues considering different rules for updating the penalty parameter.

We have also compared the behaviour of our algorithm with the sequential quadratic programming (SQP) algorithm implemented in the MatLab[1] routine `fmincon` (see e.g. [25] Chapter 18). The relative error between the multiplier obtained by the (SQP) routine and $\bar\lambda^\theta$ is of order 0.9; in fact, this routine proposes a multiplier near to the dual solution $[2\lambda^\theta,0]$. This illustrates the convergence of our method to a specific multiplier, namely the θ^*-center.

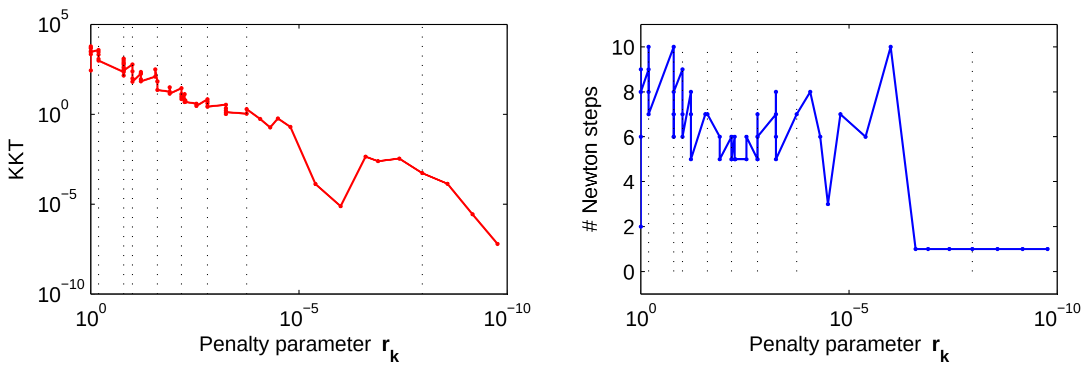

Figure 2: Evolution of the KKT residues and the number of Newton steps in (33) with respect to r_k, with vertical dotted lines each 10 main iterations.

We conclude these illustrations with Figure 2, which shows the evolution of the value of the stopping KKT criterion (35), and the number of Newton's steps in the inner iterations in (33), with respect to the value of the parameter r_k updated through the feedback rule (34). As the results show, the number of Newton's steps does not increase too much when the parameter r_k goes to 0,

[1] Matlab is registered trademarks of The MathWorks, Inc.

which seems to show that the active set identification we propose does work well when the feedback rule is used. This issue, as well as a serious study of the speed of convergence of the algorithm proposed in §5.1, is out of the scope of this paper, but shall be the subject of a future research.

References

[1] W. Achtziger. Topology optimization of discrete structures: an introduction in view of computational and nonsmooth aspects. In *Topology optimization in structural mechanics*, volume 374 of *CISM Courses and Lectures*, pages 57–100. Springer, Vienna, 1997.

[2] W. Achtziger, M. Bendsøe, A. Ben-Tal, and J. Zowe. Equivalent displacement based formulations for maximum strength truss topology design. *Impact Comput. Sci. Engrg.*, 4(4):315–345, 1992.

[3] F. Alvarez and M. Carrasco. Minimization of the expected compliance as an alternative approach to multiload truss optimization. *Struct. Multidiscip. Optim.*, 29(6):470–476, 2005.

[4] F. Alvarez, M. Carrasco, and K. Pichard. Convergence of a hybrid projection-proximal point algorithm coupled with approximation methods in convex optimization. *Mathematics of Operations Research*, 30(4):966–984, 2005.

[5] F. Alvarez and R. Cominetti. Primal and dual convergence of a proximal point exponential penalty method for linear programming. *Mathematical Programming*, 93(1, Ser. A):87–96, 2002.

[6] H. Attouch. Viscosity solutions of minimization problems. *SIAM J. of Optim.*, 6(3):769–806, 1996.

[7] A. Auslender, R. Cominetti, and M. Haddou. Asymptotic analysis for penalty methods in convex and linear programming. *Mathematics of Operations Research*, 22(1):43–62, 1997.

[8] A. Auslender, J.P. Crouzeix, and P. Fedit. Penalty-proximal methods in convex programming. *J. of Optimization Theory and Applications*, 55:1–21, 1987.

[9] M. P. Bendsøe and O. Sigmund. *Topology optimization*. Springer-Verlag, Berlin, 2003. Theory, methods and applications.

[10] A. Brøndsted and R. T. Rockafellar. On the subdifferentiability of convex functions. *Proc. Amer. Math. Soc.*, 16:605–611, 1965.

[11] M. Carrasco. Hybrid extragradient proximal algorithm coupled with parametric approximation and penalty/barrier methods. to appear, 2007.

[12] T. Champion. Tubularity and asymptotic convergence of penalty trajectories in convex programming. *SIAM J. of Optim.*, 13(1):212–227, 2002.

[13] R. Cominetti. Coupling the proximal point algorithm with approximation methods. *J. of Optimization Theory and Applications*, 95(3):581–600, 1997.

[14] R. Cominetti. Nonlinear averages and convergence of penalty trajectories in convex programming. In *Ill-posed variational problems and regularization techniques (Trier, 1998)*, volume 477 of *Lecture Notes in Econom. and Math. Systems*, pages 65–78. Springer, Berlin, 1999.

[15] R. Cominetti and M. Courdurier. Coupling general penalty schemes for convex programming with the steepest descent and the proximal point algorithm. *SIAM J. of Optim.*, 13(3):745–765 (electronic) (2003), 2002.

[16] R. Cominetti and J.-P. Dussault. Stable exponential-penalty algorithm with superlinear convergence. *J. Optim. Theory Appl.*, 83(2):285–309, 1994.

[17] R. Cominetti and J. M. Pérez-Cerda. Quadratic rate of convergence for a primal-dual exponential penalty algorithm. *Optimization*, 39(1):13–32, 1997.

[18] R. Cominetti and J. San Martín. Trajectory analysis for the exponential penalty method in linear programming. *Mathematical Programming*, 67(2, Ser. A):169–187, 1994.

[19] Clovis C. Gonzaga and Rómulo A. Castillo. A nonlinear programming algorithm based on non-coercive penalty functions. *Math. Program.*, 96(1, Ser. A):87–101, 2003.

[20] F. Jarre, M. Kočvara, and J. Zowe. Optimal truss design by interior-point methods. *SIAM J. Optim.*, 8(4):1084–1107 (electronic), 1998.

[21] A. Kaplan. On a convex programming method with internal regularization. *Soviet Mathematics Doklady*, 19:795–799, 1978.

[22] A. Kaplan and R. Tichatschke. Proximal point methods in view of interior-point strategies. *J. of Optimization Theory and Applications*, 98(2):399–429, 1998.

[23] I. V. Konnov. Combined relaxation methods for the search for equilibrium points and solutions of related problems. *Izv. Vyssh. Uchebn. Zaved. Mat.*, (2):46–53, 1993.

[24] K. Marti and G. Stöckl. Topology and geometry optimization under stochastic uncertainty. *In: Numerical Modelling of Uncertainties (eds. Level, P. et al.), Valenciennes: Presses Universitaires de Valenciennes*, pages 55–63, 2000.

[25] J. Nocedal and S.J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.

[26] M.V. Solodov and B.F. Svaiter. A hybrid projection-proximal point algorithm. *J. of Convex Analysis*, 6(1):59–70, 1999.