Convergence of a hybrid projection-proximal point algorithm coupled with approximation methods in convex optimization

Felipe Alvarez
Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (CNRS UMI 2807), Universidad de Chile, Casilla 170/3, Correo 3, Santiago, Chile.
email: falvarez@dim.uchile.cl

Miguel Carrasco
Departamento de Ingeniería Matemática, Universidad de Chile, Casilla 170/3, Correo 3, Santiago, Chile.
email: migucarr@dim.uchile.cl

Karine Pichard
Centro de Modelamiento Matemático (CNRS UMI 2807), Universidad de Chile, Casilla 170/3, Correo 3, Santiago, Chile.
email: kpichard@dim.uchile.cl

In order to minimize a closed convex function that is approximated by a sequence of better behaved functions, we investigate the global convergence of a general hybrid iterative algorithm, which consists of an inexact relaxed proximal point step followed by a suitable orthogonal projection onto a hyperplane. The latter permits to consider a fixed relative error criterion for the proximal step. We provide various sets of conditions ensuring the global convergence of this algorithm. The analysis is valid for nonsmooth data in infinite-dimensional Hilbert spaces. Some examples are presented, focusing on penalty/barrier methods in convex programming. We also show that some results can be adapted to the zero-finding problem for a maximal monotone operator.

Key words: Parametric approximation; diagonal iteration; proximal point; hybrid method ; global convergence

MSC2000 Subject Classification: Primary: 90C25, 65K05; Secondary: 49M30, 90C51

OR/MS subject classification: Primary: Programming; Secondary: Nonlinear, Algorithms

---

1. Introduction Throughout this paper, H stands for a real Hilbert space. The scalar product and norm in H are respectively denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$. Let $\Gamma_0(H)$ be the class of all the extended-real-valued functions $f\colon H \to \mathbb{R} \cup \{\infty\}$ such that f is l.s.c. (lower semicontinuous), proper (i.e., $f \not\equiv \infty$) and convex. If $f \in \Gamma_0(H)$, the effective domain of f is given by $\operatorname{dom} f = \{x \in H \mid f(x) < \infty\}$, and the δ-subdifferential[1] of f at x is defined by $\partial_\delta f(x) := \{g \in H \mid \forall y \in H,\ f(x) + \langle g, y - x\rangle \le f(y) + \delta\}$, $\delta \ge 0$.

Let $\overline{f} \in \Gamma_0(H)$ and suppose that $\operatorname{Argmin} \overline{f}$, the set of all the minimizers of $\overline{f}$, is nonempty. Given a sequence $(f_k)_{k\in\mathbb{N}} \subset \Gamma_0(H)$ of functions converging, in a sense to be made precise later, to $\overline{f}$ as $k \to \infty$, we consider the sequences $(x^k)_{k\in\mathbb{N}} \subset H$ which are generated by the following Diagonal Hybrid Projection-Proximal Point Algorithm (DHP-PPA):

- **Proximal step.** Given $x^k \in H$, $\lambda_k > 0$, $\delta_k \ge 0$ and $\rho_k \in (0, 2)$, find $z^k \in H$ such that
$$(z^k - x^k)/\lambda_k + g^k = \xi^k, \text{ for } g^k \in \rho_k \partial_{\delta_k} f_k(z^k_\rho), \tag{1}$$
where
$$z^k_\rho = z^k/\rho_k + (1 - 1/\rho_k)x^k, \tag{2}$$
and the residue $\xi^k \in H$ is required to satisfy the following condition:
$$\|\xi^k\| \le \sigma \sqrt{\|z^k - x^k\|^2/\lambda_k^2 + \|g^k\|^2}, \tag{3}$$
where $\sigma \in [0, 1)$ is a fixed relative error tolerance.
- **Projection step.** If $g^k = 0$ then set $x^{k+1} = x^k$; otherwise, take
$$x^{k+1} = x^k - \beta_k g^k, \text{ with } \beta_k = \langle g^k, x^k - z^k\rangle / \left\|g^k\right\|^2. \tag{4}$$
Let $k \leftarrow k + 1$ and return to the proximal step.

The algorithm could be called an "inexact" HPPA, but we prefer to name this algorithm "diagonal" because a single hybrid projection-proximal iteration is applied to f_k and then the objective function is

[1] See [36] for further details on the approximate subdifferential.

updated to f_{k+1}. Since $f_k \to \overline{f}$ as $k \to \infty$, this procedure is expected to approach the set $\operatorname{Argmin} \overline{f}$. In this paper, we focus our attention on obtaining general conditions ensuring the convergence of $(x^k)_{k\in\mathbb{N}}$ towards a minimizer of $\overline{f}$, under the following hypotheses on the parameters:

$$\sum \lambda_k \delta_k < \infty, \tag{5}$$

$$R_1 := \inf_{k\geq 0} \rho_k > 0 \text{ and } R_2 := \sup_{k\geq 0} \rho_k < 2. \tag{6}$$

In order to motivate the DHP-PPA, we begin by noticing that if $\delta_k = 0$ then (1) and (2) amount to

$$z^k = (1-\rho_k)x^k + \rho_k J_{\lambda_k}^{\partial f_k}(x^k + (\lambda_k/\rho_k)\xi^k), \tag{7}$$

where the single-valued function $J_\lambda^{\partial f_k} := (I+\lambda\partial f_k)^{-1} : H \to H$ is the resolvent of ∂f_k of parameter λ [17]. If in addition $\xi^k = 0$ then, by (4), $x^{k+1} = z^k$, and it follows from (7) that $x^{k+1} = (1-\rho_k)x^k + \rho_k J_{\lambda_k}^{\partial f_k}(x^k)$. Taking $\rho_k \equiv 1$ we recover an exact iteration of the standard PPA, introduced in [44] for solving some variational inequalities. When $\rho_k \in (0,2)$, this becomes an iteration of the relaxed PPA which was introduced in [31] (see also [32] and [33] for more details) for speeding up convergence; see [15, pp. 129-131] and [26] for some illustrations of such an accelerating effect in the case of over-relaxation, that is, when $\rho_k \in (1,2)$. In the case where the approximate subdifferential $\partial_{\delta_k} f_k$ is replaced by $A : H \rightrightarrows H$, a fixed maximal monotone operator (see [17] and §5), global convergence of the standard PPA towards a solution to $0 \in A(x)$ was established in [53], permitting some inexact iterations under summability conditions on the errors. Similar results were obtained in [25] for the relaxed PPA. Specific convergence results for convex minimization, where one is interested in solving the stationary point condition $0 \in \partial\overline{f}(x)$, were investigated in [24, 35, 54]. In this minimization context, finite bundle methods to find approximate PPA iterates have been studied for nonsmooth data [9, 16, 36].

On the other hand, (4) is a projection step because it can be written as $x^{k+1} = P_k x^k$, where $P_k \colon H \to H$ is the orthogonal projection operator onto the hyperplane $\{x \in H \mid \langle g^k, x - z^k\rangle = 0\}$. By monotonicity and Lemma 2.2(ii) below, the latter separates the current iterate x^k from the stationary set $S_k = \{x \in H \mid 0 \in \partial f_k(x)\}$. Thus, in this algorithm the proximal iteration is used to construct this separating hyperplane, the next iterate x^{k+1} is then obtained by a trivial projection of x^k, which is not expensive at all from a numerical point of view. The hyperplane projection method was first proposed in [40] with the name "combined relaxation"; see also [41] and [28, Chap.12] for more details. Taking $\delta_k = 0$ and $\rho_k = 1$, (1)-(4) corresponds to one iteration of the Hybrid Projection-Proximal Point Algorithm (HP-PPA) proposed in [55] for the maximal and monotone inclusion problem $0 \in A(x)$. It is shown in [55] that HP-PPA has the remarkable property of permitting the fixed relative error tolerance $\|\xi^k\| \leq \sigma \max\{\|z^k - x^k\|/\lambda_k, \|g^k\|\}$, a less stringent condition than summability of errors, without affecting the global convergence of the algorithm. This result was improved in [56] by considering (3) as the error tolerance. Under such fixed relative error tolerances, the hyperplane projection is in general necessary to ensure the boundedness of the iterates [55, p. 62], even for minimization problems in which $A = \partial\overline{f}$ [30].

In [3] it is shown that some accelerating techniques, including relaxation, can be combined with HP-PPA iterations to obtain a globally convergent scheme for the inclusion problem $0 \in A(x)$. In particular, the results of [3, 55] can be used to solve directly the stationary point condition $0 \in \partial\overline{f}(x)$. However, when $\overline{f}$ is not strongly convex, when it has an irregular behavior due to nonsmooth data, or when there are implicit constraints in its definition, it is a common practice to approximate $\overline{f}$ by a sequence $(f_k)_{k\in\mathbb{N}} \subset \Gamma_0(H)$ of better behaved functions (e.g. viscosity methods, Tikhonov's regularization, smoothing techniques, penalty/barrier methods). Relying on such an approximating sequence, diagonal algorithms perform a prescribed number of iterations of an optimization method applied to f_k and then update the objective function to f_{k+1}, which is expected to be closer to $\overline{f}$. This is the case of (1)-(4), where a single iteration of the HP-PPA is applied to f_k, then we continue with f_{k+1}.

The diagonal approach has already been considered by several authors through purely proximal iterations. The first work in this direction seems to be [38], where the author investigates the combination of the PPA iteration with a class of interior penalties; see [39] for more recent results. See [11] for some exterior penalties, and [1, 12, 42, 43] for extensions via variational convergence methods. See [47] for different regularization-penalty methods, and [50] for some results on Tikhonov's regularization. A two-parameter exponential penalty-PPA for convex programs can be found in [48, 49]. See [20] for general results exploiting the existence of "central/optimal paths" together with applications to the log-barrier and the exponential penalty in linear programming. For improvements, extensions and primal-dual convergence results, see [4, 22]. Although some of these works consider a residual error ξ^k as in (1), they all require the sequence of errors to satisfy a summability condition (cf. (20)), a rather restrictive hypothesis for practical implementations. As already mentioned, the advantage of DHP-PPA is the fixed relative error criterium (3).

This paper is organized as follows. In §2, we discuss the well-definiteness of the sequences $(x^k)_{k\in\mathbb{N}}$ generated by (1)-(4) and establish some preliminary lemmas. In §3, we prove a general convergence result, whose potential applications as well as its drawbacks are illustrated through some examples. In §4, we investigated the special case of one-parameter approximation schemes, providing convergence results under "fast" or "slow" parametrization conditions that complete and improve the general result of §3; examples are given for penalty/barrier methods in convex programming. Finally, in §5, we briefly discuss some extensions to the zero-finding problem for maximal monotone operators.

2. Preliminaries Consider a family of functions $(f_k)_{k\in\mathbb{N}} \subset \Gamma_0(H)$. Let us begin with a brief discussion on the well-definiteness of the sequences $(x^k)_{k\in\mathbb{N}}$ generated by (1)-(4).

Given $k \geq 0$, let us introduce the auxiliary objective function

$$\phi_k(x) = \tfrac{1}{2\lambda_k}\|x - x^k\|^2 + \rho_k^2 f_k(x/\rho_k + (1 - 1/\rho_k)x^k),$$

which is strongly convex and coercive. Therefore, ϕ_k admits a unique global minimizer, which we denote by y^k and is characterized by the stationary condition $0 \in \partial\phi_k(y^k)$. The latter amounts to

$$(y^k - x^k)/\lambda_k + g^k = 0,$$

for some $g^k \in \rho_k\partial f_k(y^k_\rho)$ where $y^k_\rho = y^k/\rho_k + (1 - 1/\rho_k)x^k$.

This shows that the algorithm (1)-(4) is well defined in the sense that there exists z^k satisfying (1)-(3) (it suffices to take $z^k = y^k$, $\xi^k = 0$ and any $\delta_k \geq 0$). Nevertheless, since the exact minimization of ϕ_k cannot be attained in practice, inexact computations are essential for implementable versions of the algorithm. In this direction, the choice of a specific method for solving the inexact proximal subproblem should depend on the data regularity. For instance, if f_k is locally Lipschitz continuous then this can be done in a finite number of operations for a given $\delta_k > 0$ and $\xi^k = 0$ by means of a bundle method (see [9, 24]). The total number of operations can be bounded from above in terms of the algorithm parameters, some appropriate local estimates for the Lipschitz constant and a factor of the form $1/\delta_k^2$. When f_k is $C^{1,1}$, this factor can be improved to $1/\delta_k$; see [20, Appendix B] for all details.

In the case of smooth data, any standard descent method (steepest descent, conjugate gradient, Newton, BFGS,...) for the approximate minimization of ϕ_k is able to find, in a finite number of iterations, points such that the norm of the gradient $\nabla\phi_k$ is as small as the machine precision allows one to do. Thus, in an implementable version of the algorithm, z^k will be found by applying a smooth descent method with $\delta_k = 0$ and $\xi^k = \nabla\phi_k(z^k)$, and under (3) as the stopping rule. In this case, the computational complexity of the subproblem algorithm in terms of the total number of iterations is a very interesting issue which is beyond the scope of this paper and it will not be treated here.

Notice that, even when the approximate f_k is smooth, the use of PPA may be interesting because of the eventual ill-conditioning of f_k as it approaches the limit function $\overline{f}$, which may have multiple minimizers and realize the value ∞. In fact, the multiplicity of optimal solutions usually leads to a poorly scaled approximate f_k, and the descent direction may not provide much reduction in the function when computed directly from f_k. This forces any line search method to choose a very small step length in order to avoid zigzagging iterates and drastic increases in numerical instabilities, impairing the overall efficiency of the algorithm. On the other hand, the quadratic term in ϕ_k acts as a stabilizing technique. In particular, if f_k is smooth then the Hessian $\nabla^2\phi_k = \frac{1}{\lambda_k}I + \nabla^2 f_k$ is always positive definite, whereas $\nabla^2 f_k$ may be degenerate.

From now on, we turn our attention to the global convergence of the sequence $(x^k)_{k\in\mathbb{N}}$. To this end, we conclude this preliminary section by proving two elementary results that will be useful for the sequel. First, let us recall a result of [56] concerning the relative error criterion (3).

LEMMA 2.1 *Let $\sigma \in [0,1)$ and define $\varrho = \sqrt{1-(1-\sigma^2)^2}$. If $v = u + \xi$ where $\|\xi\|^2 \leq \sigma^2(\|u\|^2 + \|v\|^2)$ then:*

(*i*) $\langle u,v\rangle \geq (\|u\|^2+\|v\|^2)(1-\sigma^2)/2$.

(*ii*) $(1-\varrho)\|v\| \leq \|u\|(1-\sigma^2) \leq (1+\varrho)\|v\|$.

Proof. (i) is immediate from $\|\xi\|^2 = \|u\|^2 - 2\langle u,v\rangle + \|v\|^2$. Applying the Cauchy-Schwarz inequality to (i), we get $Q(\|u\|)\leq 0$ for $Q(z) = z^2 - \frac{2}{1-\sigma^2}z\|v\| + \|v\|^2$, whose roots are $z_\pm = (1\pm\varrho)\|v\|/(1-\sigma^2)$. This proves (ii). □

Let us now state some useful estimates which are direct consequences of the previous lemma.

Lemma 2.2 *Under* (1)-(4), *if* $\varrho = \sqrt{1-(1-\sigma^2)^2}$ *then we have:*

(*i*) $\lambda_k(1-\varrho)\|g^k\| \leq (1-\sigma^2)\|x^k - z^k\| \leq \lambda_k(1+\varrho)\|g^k\|$.

(*ii*) $\lambda_k(1-\sigma^2)/2\left(\|g^k\|^2 + \|x^k-z^k\|^2/\lambda_k^2\right) \leq \langle g^k, x^k - z^k\rangle$.

(*iii*) $\beta_k \in [\lambda_k(1-\sigma^2)/2, \lambda_k(1+\varrho)/(1-\sigma^2)]$.

Proof. We apply Lemma 2.1 to $v = g^k$, $u = (x^k - z^k)/\lambda^k$ and $\eta = \xi$ to get (i) and (ii). For (iii), using first the Cauchy-Schwarz inequality and then (i), we get $\beta_k \leq \|x^k - z^k\|/\|g^k\| \leq \lambda_k(1+\varrho)/(1-\sigma^2)$. On the other hand, (ii) implies that $\beta_k \geq \lambda_k(1-\sigma^2)/2(1+\|x^k-z^k\|^2/(\|g^k\|^2\lambda_k^2))$, this leads to (iii). □

Remark 2.1 Suppose that $g^k = 0$. As $0 \in \partial_{\delta_k} f_k(z_\rho^k)$, this implies that z_ρ^k is a δ_k-minimizer of f_k. Moreover, by virtue of Lemma 2.2(i), $z^k - x^k = 0$, hence $\xi^k = 0$ and $z_\rho^k = z^k = x^k$. Therefore, in that case, $x^{k+1} = x^k$ is a δ_k-minimizer of f_k. On the other hand, assuming $g^k \neq 0$, Lemma 2.2 yields $\langle g^k, x^k - z^k\rangle > 0$. As $g^k \in \rho_k\partial_{\delta_k} f_k(z_\rho^k)$, it is easy to see that for all $\bar{x} \in \operatorname{Argmin} f_k$, $\langle g^k, \bar{x} - z^k\rangle \leq (1-1/\rho_k)\langle g^k, x^k - z^k\rangle + \rho_k\delta_k$. Thus, if $\delta_k < \langle g^k, x^k - z^k\rangle/\rho_k^2$ then the hyperplane $\{x \in H \mid \langle g^k, x - z^k\rangle = \alpha\}$ strictly separates x^k from $\operatorname{Argmin} f_k$ whenever α satisfies $(1-1/\rho_k)\langle g^k, x^k-z^k\rangle + \rho_k\delta_k \leq \alpha \leq \langle g^k, x^k - z^k\rangle$. The latter is the geometric motivation for the projection step (4).

Finally, all our convergence results rely strongly on the elementary identity that is established in the following result.

Lemma 2.3 *Under* (1)-(4), *for any* $u \in H$, *we have*

$$\|x^{k+1} - u\|^2 = \|x^k - u\|^2 + 2\beta_k\langle g^k, u - z_\rho^k\rangle + (1-2/\rho_k)\beta_k^2\|g^k\|^2.$$

Proof. By direct computations, $\|x^{k+1}-u\|^2 = \|x^k - u\|^2 + 2\langle x^{k+1} - x^k, x^k - u\rangle + \|x^{k+1} - x^k\|^2 = \|x^k - u\|^2 - 2\beta_k\langle g^k, x^k - u\rangle + \beta_k^2\|g^k\|^2$. But $x^k = z_\rho^k + (x^k - z^k)/\rho_k$, hence $\langle g^k, x^k - u\rangle = \langle g^k, z_\rho^k - u\rangle + \langle g^k, x^k - z^k\rangle/\rho_k = \langle g^k, z_\rho^k - u\rangle + \beta_k\|g^k\|^2/\rho_k$, and the proof is complete. □

3. A first general result on global convergence Let $\overline{f} \in \Gamma_0(H)$. In our first convergence result, we suppose that $\overline{f}$ and the sequence of functions $(f_k)_{k\in\mathbb{N}} \subset \Gamma_0(H)$ satisfy the following conditions:

$$\operatorname{Argmin}\overline{f} \neq \emptyset. \tag{8}$$

$$\forall k\in\mathbb{N},\ \forall x\in H,\ \overline{f}(x) \leq f_k(x). \tag{9}$$

$$\forall k\in\mathbb{N},\ \forall\bar{x}\in\operatorname{Argmin}\overline{f},\ \exists\theta_k(\bar{x})\geq 0\text{: } f_k(\bar{x}) \leq \min\overline{f} + \theta_k(\bar{x}). \tag{10}$$

Remark 3.1 Similar conditions are considered in [12, 43] for some diagonal proximal methods.

Theorem 3.1 *Let* $(x^k)_{k\in\mathbb{N}}$ *be a sequence generated by* (1)-(4) *under* (5) *and* (6). *If* (8), (9) *and* (10) *hold, and if we assume that*

$$\forall\bar{x}\in\operatorname{Argmin}\overline{f},\ \sum\lambda_k\theta_k(\bar{x}) < \infty, \tag{11}$$

then:

(i) For every $\bar{x} \in \operatorname{Argmin} \overline{f}$*, the sequence* $\left(\|x^k - \bar{x}\|\right)_{k\in\mathbb{N}}$ *converges and we also have:*

$$\sum \lambda_k^2 \left\|g^k\right\|^2 < \infty, \; \sum \left\|z^k - x^k\right\|^2 < \infty \;\; \textit{and} \; \sum \lambda_k[\overline{f}(z_\rho^k) - \min \overline{f}] < \infty.$$

(ii) If $\dim H < \infty$ *and* $\sum \lambda_k = \infty$ *then* $(x^k)_{k\in\mathbb{N}}$ *converges to a point in* $\operatorname{Argmin} \overline{f}$*.*

(iii) If $\dim H = \infty$ *and* $\inf_k \lambda_k > 0$ *then* $(x^k)_{k\in\mathbb{N}}$ *weakly converges to a point in* $\operatorname{Argmin} \overline{f}$*.*

Proof. (i) Let $\bar{x}$ be in $\operatorname{Argmin} \overline{f}$ and define $\varphi_k = \left\|x^k - \bar{x}\right\|^2$. Applying Lemma 2.3 to $u = \bar{x}$, and since $g^k/\rho_k \in \partial_{\delta_k} f_k(z_\rho^k)$, we deduce that

$$\varphi_{k+1} \le \varphi_k + 2\rho_k\beta_k[f_k(\bar{x}) - f_k(z_\rho^k) + \delta_k] + (1 - 2/\rho_k)\beta_k^2 \left\|g^k\right\|^2. \tag{12}$$

Using (9) with $x = z_\rho^k$ we obtain

$$\varphi_{k+1} \le \varphi_k + 2\rho_k\beta_k[f_k(\bar{x}) - \overline{f}(z_\rho^k)] + 2\rho_k\beta_k\delta_k + (1 - 2/\rho_k)\beta_k^2 \left\|g^k\right\|^2. \tag{13}$$

By (10), we deduce that

$$\varphi_{k+1} \le \varphi_k + 2\rho_k\beta_k[\overline{f}(\bar{x}) - \overline{f}(z_\rho^k)] + 2\rho_k\beta_k[\theta_k(\bar{x}) + \delta_k] + (1 - 2/\rho_k)\beta_k^2 \left\|g^k\right\|^2. \tag{14}$$

The optimality of $\bar{x}$ yields $\overline{f}(\bar{x}) \le \overline{f}(z_\rho^k)$ and, consequently,

$$\varphi_{k+1} \le \varphi_k + 2\rho_k\beta_k[\theta_k(\bar{x}) + \delta_k] + (1 - 2/\rho_k)\beta_k^2 \left\|g^k\right\|^2. \tag{15}$$

Since $\rho_k \le R_2$, we have $\varphi_{k+1} \le \varphi_k + 2R_2\beta_k[\theta_k(\bar{x}) + \delta_k]$. By Lemma 2.2, we obtain the following estimate:

$$\varphi_{k+1} \le \varphi_k + \zeta_k,$$

for

$$\zeta_k = R_2\lambda_k(1 + \varrho)[\theta_k(\bar{x}) + \delta_k]/(1 - \sigma^2).$$

Assumptions (5) and (11) ensure that $\sum \zeta_k < \infty$. We deduce that $(\varphi_k + \sum_{j\ge k} \zeta_j)_{k\in\mathbb{N}}$ is nonincreasing, hence $(\varphi_k)_{k\in\mathbb{N}}$ is convergent.

Let us return to (15). Using the fact that $0 < R_1 \le \rho_k \le R_2 < 2$ and Lemma 2.2, we have

$$\frac{(2 - R_2)(1 - \sigma^2)^2}{4R_2}\lambda_k^2 \left\|g^k\right\|^2 \le \varphi_k - \varphi_{k+1} + \zeta_k. \tag{16}$$

Summing over k, we obtain

$$\frac{(2 - R_2)(1 - \sigma^2)^2}{4R_2}\sum \lambda_k^2 \left\|g^k\right\|^2 \le \varphi_0 - \varphi_\infty + \sum \zeta_k < \infty, \tag{17}$$

where φ_∞ is the limit of $(\varphi_k)_{k\in\mathbb{N}}$. It follows that $\sum \lambda_k^2 \left\|g^k\right\|^2 < \infty$. Moreover, by Lemma 2.2(i), we also obtain $\sum \left\|z^k - x^k\right\|^2 < \infty$. We proceed analogously, summing over k in (12), to deduce that

$$R_1(1 - \sigma^2)\sum \lambda_k[\overline{f}(z_\rho^k) - \overline{f}(\bar{x})] \le \varphi_0 - \varphi_\infty + \sum \zeta_k < \infty. \tag{18}$$

Consequently,

$$\sum \lambda_k[\overline{f}(z_\rho^k) - \overline{f}(\bar{x})] < \infty, \tag{19}$$

and the proof of (i) is complete.

(ii) Under the assumption $\sum \lambda_k = \infty$, a consequence of (19) is the following:

$$\liminf_{k\to\infty} \overline{f}(z_\rho^k) = \overline{f}(\bar{x}).$$

Since $(x^k)_{k\in\mathbb{N}}$ is bounded by (i), we can extract a convergent subsequence $(x^{k_j})_{j\in\mathbb{N}}$ whose limit is denoted by $\hat{x}$. But $\lim_{k\to\infty} \left\|z^k - x^k\right\| = 0$, hence $(z_\rho^{k_j})_{j\in\mathbb{N}}$ also converges to $\hat{x}$. By the lower semi-continuity of $\overline{f}$, we have

$$\overline{f}(\hat{x}) \le \lim_{j\to\infty} \overline{f}(z_\rho^{k_j}) = \liminf_{k\to\infty} \overline{f}(z_\rho^k) = \overline{f}(\bar{x}).$$

From the optimality of $\bar{x}$, it is clear that $\hat{x}$ belongs to $\operatorname{Argmin} \overline{f}$. Finally, we use (i) to conclude that $\lim_k \left\|x^k - \hat{x}\right\| = \lim_{j\to\infty} \left\|x^{k_j} - \hat{x}\right\| = 0$ so that all the sequence $(x^k)_{k\in\mathbb{N}}$ converges to $\hat{x}$.

(iii) When $\dim H = \infty$, the following classical result from [52] provides a useful criterion for weak convergence without the knowledge *a priori* of the limit point.

LEMMA 3.1 (OPIAL) *Let H be a Hilbert space and $(x^k)_{k\in\mathbb{N}}$ a sequence in H such that there exists a nonempty set $C \subset H$ satisfying:*

(a) For every $\bar{x} \in C$, $\lim_k \|x^k - \bar{x}\|$ exists.

(b) If $(x^{k_j})_{j\in\mathbb{N}}$ weakly converges to $\hat{x}$ then $\hat{x} \in C$.

Then, there exists $x^\infty \in C$ such that $(x^k)_{k\in\mathbb{N}}$ weakly converges to x^∞.

By (i), the first condition of Opial's lemma holds for $C = \operatorname{Argmin} \overline{f}$. Now, let $(x^{k_j})_{j\in\mathbb{N}}$ be a subsequence of $(x^k)_{k\in\mathbb{N}}$ weakly converging to a point $\hat{x}$. We are going to prove that $\hat{x}$ belongs to $\operatorname{Argmin} \overline{f}$. Since $\inf_k \lambda_k > 0$, (19) leads to $\lim_k \overline{f}(z^k_\rho) = \overline{f}(\bar{x}) = \min \overline{f}$. By convexity and lower semi-continuity of $\overline{f}$, we obtain that $f(\hat{x}) = \min \overline{f}$, which proves that the second condition of Opial's lemma is fulfilled and the conclusion follows. □

REMARK 3.2 Of course, Theorem 3.1 can be applied to the constant sequence $f_k \equiv \overline{f}$ with $\theta_k \equiv 0$ so that (11) is automatically satisfied. If $\delta_k \equiv 0$, this case is recovered by applying the results of [3, 55] to $A = \partial \overline{f}$.

REMARK 3.3 It follows from Theorem 3.1(i) that, *a posteriori*, the sequence of residues (ξ^k) satisfies

$$\sum \lambda_k^2 \|\xi^k\|^2 < \infty.$$

However, it may occur that $\sum \lambda_k \|\xi^k\| = \infty$; see [30] for an example based on [35] with $\rho_k \equiv 1$. The constant relative error criterion (3) is thus less stringent than the requirement

$$\sum \lambda_k \|\xi^k\| < \infty. \tag{20}$$

On the other hand, (20) is in general sufficient for the global convergence of diagonal purely proximal algorithms (see, for instance, [4, 22] and the references therein). A natural question is whether a purely proximal iteration is enough in order to ensure convergence under the fixed relative error tolerance, without assuming (20) of course. The answer turns out to be negative: the hyperplane projection is in general necessary to ensure the boundedness of the iterates (see [55, p. 62] and [30]).

EXAMPLE 3.1 *Viscosity method.* Let $f \in \Gamma_0(H)$ and assume that $\operatorname{Argmin} f$ is nonempty. Given $h \in \Gamma_0(H)$, a strictly convex, continuous, coercive and finite-valued function with $\inf_H h = 0$, and $\varepsilon > 0$, a small parameter intended to go to 0, the viscosity (or generalized Tikhonov) method [5, 13] consists in approximating f by

$$f(x, \varepsilon) = f(x) + \varepsilon h(x).$$

The special case where $h(x) = \frac{1}{2}\|x\|^2$ is referred to as the *Tikhonov regularization* of f. Given a sequence $(\varepsilon_k)_{k\in\mathbb{N}}$ with $\varepsilon_k \to 0$ as $k \to \infty$, setting

$$f_k(x) = f(x) + \varepsilon_k h(x),$$

it follows that (9) and (10) hold with $\overline{f} = f$ and $\theta_k(x) = \varepsilon_k h(x)$. In order to have (11), it suffices to request that

$$\sum \varepsilon_k \lambda_k < \infty, \tag{21}$$

and this is necessary for (11) whenever $h(\bar{x}) > 0$ for some $\bar{x} \in \operatorname{Argmin} f$. □

EXAMPLE 3.2 *Log-exp approximation of minimax problems.* Let $h_i \in \Gamma_0(H)$, $i = 1, \ldots, m$ for $m \geq 2$, and define

$$F(x) := \max_{1\leq i\leq m} \{h_i(x)\} \in \Gamma_0(H).$$

We are interested in finding a point in $\operatorname{Argmin} F$, which is supposed to be nonempty. In general, due to the max operation, F is not smooth even if every h_i is so, and this feature is an inconvenient for the direct application to F of an optimization algorithm. One standard way to regularize F is to consider the log-exp approximation [14]:

$$f(x, \varepsilon) = \varepsilon \log \left(\sum_{i=1}^{m} \exp[h_i(x)/\varepsilon] \right),$$

which preserves the regularity of the data and satisfies $F(x) \le f(x,\varepsilon) \le F(x) + \varepsilon \log m$ for all $x \in H$ and $\varepsilon > 0$. If we take $\varepsilon_k \to 0$ and define $f_k(x) = f(x, \varepsilon_k)$ then (9) and (10) are satisfied for $\overline{f} = F$ and $\theta_k \equiv \varepsilon_k \log m$. Hence, in this special case (11) is again equivalent to (21). □

EXAMPLE 3.3 *Penalty/barrier methods in convex programming.* Let us turn our attention to a mathematical program of the type

$$\min_{x \in H}\{f(x) \mid h_i(x) \le 0, i = 1, \dots, m\}, \tag{22}$$

where $f \in \Gamma_0(H)$ and $h_i \in \Gamma_0(H)$, $i = 1, \dots, m$. We assume that the set of optimal solutions of (22) is nonempty. Generally speaking, a penalty method [10, 14, 22, 27] approximates (22) with a parametric family of minimization problems of the form

$$\min_{x \in H} f(x,\varepsilon) := f(x) + \varepsilon \sum_{i=1}^{m} \psi[h_i(x)/\varepsilon], \quad \varepsilon > 0, \tag{23}$$

where the function $\psi \in \Gamma_0(\mathbb{R})$ satisfies:

(i) $(-\infty, 0) \subset \operatorname{dom}\psi$ and $\kappa := \sup(\operatorname{dom}\psi) \ge 0$.

(ii) $\psi : (-\infty, \kappa) \to \mathbb{R}$ is differentiable, increasing and strictly convex.

(iii) $\lim_{s \to -\infty} \psi'(s) = 0$ and $\lim_{s \to \kappa} \psi'(s) = \infty$.

Two distinguished examples are the exponential penalty $\psi_1(s) = \exp(s)$ (with $\kappa = \infty$), and the log-barrier $\psi_2(s) = -\log(-s)$ if $s < 0$ and ∞ otherwise (with $\kappa = 0$). In this framework, it seems natural to take

$$\overline{f}(x) = \begin{cases} f(x) & \text{if } h_i(x) \le 0, \ i = 1, \dots, m, \\ \infty & \text{otherwise,} \end{cases}$$

and $f_k(x) = f(x, \varepsilon_k)$ for $\varepsilon_k \to 0$. However, most penalty/barrier schemes do not satisfy the hypotheses of Theorem 3.1. For instance, if ψ is a barrier function with $\psi(0) = \infty$ and (22) is such that there exists an optimal solution $\bar{x}$ satisfying $h_{i_0}(\bar{x}) = 0$ for some i_0 (which is the usual case), then $f_k(\bar{x}) = f(\bar{x}, \varepsilon_k) = \infty$; consequently, (10) cannot be satisfied. On the other hand, if ψ is such that $\kappa > 0$ then, in general, there exist some unfeasible points x for which $f_k(x) < \infty = \overline{f}(x)$, and (9) does not hold. Therefore, in order to apply the previous result, ψ is forced to comply with the requirement that $\operatorname{dom}\psi = (-\infty, 0]$, which leaves many important penalty/barrier functions out of the scope of Theorem 3.1. In Section 4, we provide conditions for the convergence of the algorithm that are satisfied for wide classes of penalty/barrier functions. □

4. Parametric approximation schemes

4.1 Preliminary results We will improve the results of section 3 in the specific setting of parametric approximations in optimization. Let us consider a family of minimization problems of the type

$$(P^\varepsilon) \qquad v(\varepsilon) = \min\{f(x,\varepsilon) \mid x \in H\}, \quad \varepsilon > 0,$$

where for each $f(\cdot, \varepsilon) \in \Gamma_0(H)$ and define

$$S_\varepsilon := \operatorname{Argmin} f(\cdot, \varepsilon).$$

In general, $\varepsilon > 0$ is a small parameter intended to go to 0. We assume that:

$$\text{There exists a function } x : (0, \varepsilon_0] \to H \text{ such that } \forall \varepsilon \in (0, \varepsilon_0],\ x(\varepsilon) \in S_\varepsilon. \tag{24}$$

$$\text{The optimal path } x(\cdot) \text{ is absolutely continuous on } (0, \varepsilon_0] \text{ and } \int_0^{\varepsilon_0} \left| \frac{dx}{d\varepsilon} \right| d\varepsilon < \infty. \tag{25}$$

The idea of assuming the existence of an optimal path has already been exploited in the literature. This is the case of [7], [20], whose techniques are adapted here to deal with the DHP-PPA.

REMARK 4.1 The absolute continuity of $x(\cdot)$ is not difficult to establish in most applications. For instance, when $f(\cdot, \varepsilon)$ is differentiable at the minimizer $x(\varepsilon)$, we have the first order stationary condition $\nabla f(x(\varepsilon), \varepsilon) = 0$. When $\nabla^2 f(x(\varepsilon), \varepsilon)$ exists and is positive definite, the implicit function theorem ensures that the curve $\varepsilon \mapsto x(\varepsilon)$ is differentiable and satisfies $\nabla^2 f(x(\varepsilon),\varepsilon)\frac{dx}{d\varepsilon}(\varepsilon) + \frac{\partial^2 f}{\partial\varepsilon\partial x}(x(\varepsilon),\varepsilon) = 0$. On the other hand, in order to have the finite length condition in (25) it suffices that $\varepsilon \mapsto x(\varepsilon)$ be Lipschitz continuous, with a uniform upper bound on $\left|\frac{dx}{d\varepsilon}(\varepsilon)\right|$ for all $\varepsilon \in (0, \varepsilon_0]$. Although the latter is satisfied in many interesting cases (see some examples below), this is not always true. Indeed, this condition may fail for Tikhonov's regularization (see [7]).

The next result is a first step towards proving the convergence of the DHP-PPA in the context of parametric approximations.

PROPOSITION 4.1 *Assume that* (24) *and* (25) *hold. Given a sequence* $\varepsilon_k \searrow 0$ *as* $k \to \infty$, *let* $(x^k)_{k\in\mathbb{N}}$ *be a sequence generated by* (1)-(4) *applied to* $(f_k := f(\cdot,\varepsilon_k))_{k\in\mathbb{N}}$, *under* (5) *and* (6). *Then*

(*i*) *The real sequence* $(\|x^{k+1} - x(\varepsilon_k)\|)_{k\in\mathbb{N}}$ *is convergent.*

(*ii*) $\sum \lambda_k^2 \|g^k\|^2 < \infty$, $\sum \|z^k - x^k\|^2 < \infty$ *and* $\sum \lambda_k [f(z_\rho^k, \varepsilon_k) - v(\varepsilon_k)] < \infty$.

PROOF. (i) Set $f_k = f(\cdot,\varepsilon_k)$ and $\varphi_k = \|x^k - x(\varepsilon_{k-1})\|$ for all $k \in \mathbb{N}$. We have

$$\|x^k - x(\varepsilon_k)\| \le \|x^k - x(\varepsilon_{k-1})\| + \|x(\varepsilon_{k-1}) - x(\varepsilon_k)\| = \varphi_k + \zeta_k,$$

where

$$\zeta_k = \int_{\varepsilon_k}^{\varepsilon_{k-1}} \left|\frac{dx}{d\varepsilon}\right| d\varepsilon. \tag{26}$$

Since $g^k/\rho_k \in \partial_{\delta_k} f_k(z_\rho^k)$, we deduce from Lemma 2.3 applied to $u = x(\varepsilon_k)$ that φ_{k+1} satisfies

$$\varphi_{k+1}^2 \le (\varphi_k + \zeta_k)^2 + 2\rho_k\beta_k[f_k(x(\varepsilon_k)) - f_k(z_\rho^k) + \delta_k] + (1 - 2/\rho_k)\beta_k^2 \|g^k\|^2. \tag{27}$$

By the optimality of $x(\varepsilon_k)$, we obtain that $f_k(x(\varepsilon_k)) = v(\varepsilon_k) \le f_k(z_\rho^k)$. Consequently, using (6) and Lemma 2.2(iii), we get

$$\varphi_{k+1}^2 \le (\varphi_k + \zeta_k)^2 + \mu_k + (1 - 2/R_2)\beta_k^2 \|g^k\|^2, \tag{28}$$

with

$$\mu_k = 2R_2(1+\varrho)\lambda_k\delta_k/(1-\sigma^2). \tag{29}$$

Let us recall the following result on convergence of numerical sequences (see, for instance, [20]).

LEMMA 4.1 *Let* $\mu_k \ge 0$, $\zeta_k \ge 0$ *be such that* $\sum \mu_k < \infty$ *and* $\sum \zeta_k < \infty$. *If a nonnegative sequence* $(\varphi_k)_{k\in\mathbb{N}}$ *satisfies* $\varphi_{k+1}^2 \le \mu_k + (\varphi_k + \zeta_k)^2$ *for all* $k \in \mathbb{N}$, *then* φ_k *converges as* $k \to \infty$.

By (5) and (25), $\sum \mu_k < \infty$ and $\sum \zeta_k < \infty$ respectively. Since $R_2 < 2$ by (6), Lemma 4.1 allows us to conclude that $(\varphi_k)_{k\in\mathbb{N}}$ converges to a point denoted by φ_∞.

(ii) Let $M > 0$ be such that $\varphi_k \le M$ for all $k \in \mathbb{N}$. From (28), it follows that

$$(2/R_2 - 1)\beta_k^2 \|g^k\|^2 \le \varphi_k^2 - \varphi_{k+1}^2 + 2M\zeta_k + \zeta_k^2 + \mu_k.$$

Summing over k and using Lemma 2.2(iii), we obtain

$$\frac{(2-R_2)(1-\sigma^2)^2}{4R_2} \sum \lambda_k^2 \|g^k\|^2 \le \varphi_0^2 - \varphi_\infty^2 + 2M \int_0^{\varepsilon_0} \left|\frac{dx}{d\varepsilon}\right| d\varepsilon + \sum [\zeta_k^2 + \mu_k] < \infty. \tag{30}$$

Hence $\sum \lambda_k^2 \|g^k\|^2 < \infty$ From Lemma 2.2(i), it follows also $\sum \|z^k - x^k\|^2 < \infty$. Similar arguments applied to (27) lead to

$$R_1(1-\sigma^2)\sum \lambda_k [f_k(z_\rho^k) - v(\varepsilon_k)] \le \varphi_0^2 - \varphi_\infty^2 + 2M \int_0^{\varepsilon_0} \left|\frac{dx}{d\varepsilon}\right| d\varepsilon + \sum [\zeta_k^2 + \mu_k] < \infty,$$

and then the last announced result follows. □

Next, assume that the family $(f(\cdot,\varepsilon))_{\varepsilon>0}$ satisfies the following conditions:

$$\text{There exists } \overline{f} \in \Gamma_0(H) \text{ such that } \overline{f}(y^\infty) \le \liminf_{j\to\infty} f(y^j, \varepsilon_j) \text{ for every } \varepsilon_j \to 0 \text{ and } y^j \rightharpoonup y^\infty. \tag{31}$$

$$\lim_{\varepsilon\to 0} v(\varepsilon) = \min \overline{f} \text{ and there exists } x^* \in \operatorname{Argmin} \overline{f} \text{ such that } \lim_{\varepsilon\to 0} x(\varepsilon) = x^*. \tag{32}$$

REMARK 4.2 These are natural assumptions which are satisfied by large classes of approximation schemes. In particular, under apropriate conditions on the data, every example discussed in Section 3 complies with these conditions. For viscosity methods see [5, 58]; for penalty/barrier methods in linear programming, [10, 23, 45, 57]; for penalty/barrier methods in convex programming, [2, 8, 19, 21, 46].

By virtue of Proposition 4.1, an immediate consequence of these additional assumptions is the following.

PROPOSITION 4.2 *Suppose that* (31) *and* (32) *are satisfied. Under the hypotheses of Proposition* 4.1, $(x^k)_{k\in\mathbb{N}}$ *is bounded. Moreover, if* $\inf_k \lambda_k > 0$ *then* $\lim_k f(z^k_\rho, \varepsilon_k) = \min \overline{f}$ *and all the weak cluster points of* $(x^k)_{k\in\mathbb{N}}$ *belong to* $\operatorname{Argmin} \overline{f}$. *If* $\sum \lambda_k = \infty$ *then there exists at least one weak cluster point of* $(x^k)_{k\in\mathbb{N}}$ *that belongs to* $\operatorname{Argmin} \overline{f}$.

PROOF. Since $x(\varepsilon)$ converges as $\varepsilon \to 0$, it follows from Proposition 4.1(i) that $(x^k)_{k\in\mathbb{N}}$ is bounded.

Let $k_j \to \infty$ and $\hat{x}$ be such that $x^{k_j} \rightharpoonup \hat{x}$. We claim that $\hat{x} \in \operatorname{Argmin} \overline{f}$ when $\inf_k \lambda_k > 0$. Indeed, the latter together with Proposition 4.1(ii) yield $\lim_k [f(z^k_\rho, \varepsilon_k) - v(\varepsilon_k)] = 0$. By (32), we deduce that $\lim_k f(z^k_\rho, \varepsilon_k) = \min \overline{f}$. Since $\|z^k_\rho - x^k\| = |1/\rho_k - 1| \|z^k - x^k\|$, which tends to 0 due to Proposition 4.1(ii), we deduce that $z^{k_j}_\rho \rightharpoonup \hat{x}$. From (31), it follows that $\overline{f}(\hat{x}) \le \liminf_j f(z^{k_j}, \varepsilon_{k_j}) = \min \overline{f}$, which proves the optimality of $\hat{x}$.

If $\sum \lambda_k = \infty$ then Proposition 4.1(ii) implies that $\liminf_k [f(z^k_\rho, \varepsilon_k) - v(\varepsilon_k)] = 0$, hence $\lim_j f(z^{k_j}_\rho, \varepsilon_{k_j}) = \min \overline{f}$ for an appropriate subsequence $k_j \to \infty$. By similar arguments as before, up to a new subsequence, we may assume that both $(z^{k_j})_{j\in\mathbb{N}}$ and $(x^{k_j})_{j\in\mathbb{N}}$ weakly converge to some $\hat{x}$, which is a minimizer of $\overline{f}$ by virtue of (31). □

We need more assumptions to ensure the convergence of the whole sequence $(x^k)_{k\in\mathbb{N}}$ to a point in $\operatorname{Argmin} \overline{f}$. This is the goal of the next two sections.

4.2 Asymptotic convergence under fast parametrization Let us reinforce (32) by assuming that

$$\forall \varepsilon \in (0, \varepsilon_0],\ \forall \bar{x} \in \operatorname{Argmin} \overline{f}, \exists \eta > 0,\ \exists \theta(\bar{x}, \varepsilon) \ge 0\colon\ f(x(\varepsilon) + \eta(\bar{x} - x^*), \varepsilon) \le v(\varepsilon) + \theta(\bar{x}, \varepsilon), \tag{33}$$

which is a perturbed variant of (10) that was first introduced in [20].

THEOREM 4.1 *Suppose that the family* $(f(\cdot, \varepsilon))_{\varepsilon>0}$ *satisfies* (24), (25), (31), (32) *and* (33). *Let* $(x^k)_{k\in\mathbb{N}}$ *be generated by* (1)-(4) *applied to* $(f_k := f(\cdot, \varepsilon_k))_{k\in\mathbb{N}}$, *under* (5) *and* (6), *for some sequence* $\varepsilon_k \searrow 0$ *as* $k \to \infty$. *If*

$$\forall \bar{x} \in \operatorname{Argmin} \overline{f},\ \sum \lambda_k \theta(\bar{x}, \varepsilon_k) < \infty, \tag{34}$$

then:

(*i*) *For any* $\bar{x}$ *in* $\operatorname{Argmin} \overline{f}$, $\lim_k \left\|x^k - \bar{x}\right\|$ *exists.*

(*ii*) *If* $\dim H < \infty$ *and* $\sum \lambda_k = \infty$, *the sequence* $(x^k)_{k\in\mathbb{N}}$ *converges to a point in* $\operatorname{Argmin} \overline{f}$.

(*iii*) *If* $\dim H = \infty$ *and* $\inf_k \lambda_k > 0$, *the sequence* $(x^k)_{k\in\mathbb{N}}$ *weakly converges to a point in* $\operatorname{Argmin} \overline{f}$.

PROOF. (i) Let $\bar{x}$ be in $\operatorname{Argmin} \overline{f}$. We know from Proposition 4.1 that $(\varphi_k = \left\|x^k - x(\varepsilon_{k-1})\right\|)_{k\in\mathbb{N}}$ converges and then $(\left\|x^k - x^*\right\|)_{k\in\mathbb{N}}$ also converges. Since

$$\left\|x^k - \bar{x}\right\|^2 = \left\|x^k - x^*\right\|^2 + 2\langle x^k - x^*, x^* - \bar{x}\rangle + \left\|x^* - \bar{x}\right\|^2,$$

it suffices to verify that $(\langle x^k, x^* - \bar{x}\rangle)_{k\in\mathbb{N}}$ is convergent to prove (i). To this end, we begin with the following estimate:

$$\begin{aligned}\varphi_k^2 &= \varphi_{k+1}^2 + 2\langle x^{k+1} - x(\varepsilon_k), x^k - x^{k+1} - (x(\varepsilon_{k-1}) - x(\varepsilon_k))\rangle + \left\|x^k - x^{k+1} - (x(\varepsilon_{k-1}) - x(\varepsilon_k))\right\|^2 \\ &\ge \varphi_{k+1}^2 + 2\langle x^{k+1} - x(\varepsilon_k), x^k - x^{k+1}\rangle - 2\langle x^{k+1} - x(\varepsilon_k), x(\varepsilon_{k-1}) - x(\varepsilon_k)\rangle.\end{aligned}$$

As a consequence, we get

$$\langle x^{k+1} - x(\varepsilon_k), x^k - x^{k+1}\rangle \le \tfrac{1}{2}(\varphi_k^2 - \varphi_{k+1}^2) + \varphi_{k+1}\zeta_k, \tag{35}$$

where ζ_k is given by (26) and satisfies $\sum \zeta_k = \int_0^{\varepsilon_0} |dx/d\varepsilon|\, d\varepsilon < \infty$. Now we define

$$a_k := \eta\langle x^k, x^* - \bar{x}\rangle + \tfrac{1}{2}\varphi_k^2 + M\sum_{j\ge k}\zeta_j,$$

where $M > 0$ is such that $\varphi_k \le M$ for all $k \in \mathbb{N}$. We claim that $(a_k)_{k\in\mathbb{N}}$ is convergent, which proves the convergence of $(\langle x^k, x^* - \bar{x}\rangle)_{k\in\mathbb{N}}$. First, notice that $(a_k)_{k\in\mathbb{N}}$ is bounded by virtue of the convergence of $(\varphi_k)_{k\in\mathbb{N}}$. On the other hand, as Since $x^{k+1} = x^k - \beta_k g^k = z_\rho^k + (x^k - z^k)/\rho_k - \beta_k g^k$, we have that

$$\begin{aligned}
\langle x^k - x^{k+1}, x(\varepsilon_k) + \eta(\bar{x} - x^*) - x^{k+1}\rangle &= \beta_k\langle g^k, x(\varepsilon_k) + \eta(\bar{x} - x^*) - z_\rho^k\rangle + (1 - 1/\rho_k)\beta_k^2 \left\|g^k\right\|^2 \\
&\le \rho_k\beta_k\left[f_k(x(\varepsilon_k) + \eta(\bar{x} - x^*)) - f_k(z_\rho^k)\right] + \rho_k\beta_k\delta_k + \beta_k^2\left\|g^k\right\|^2 \\
&\le \rho_k\beta_k\left[f_k(x(\varepsilon_k) + \eta(\bar{x} - x^*)) - f_k(x(\varepsilon_k))\right] + \rho_k\beta_k\delta_k + \beta_k^2\left\|g^k\right\|^2 \\
&\le \rho_k\beta_k\theta(\bar{x}, \varepsilon_k) + \rho_k\beta_k\delta_k + \beta_k^2\left\|g^k\right\|^2 \\
&\le \frac{(1+\varrho)R_2}{1-\sigma^2}\lambda_k(\theta(\bar{x}, \varepsilon_k) + \delta_k) + \beta_k^2\left\|g^k\right\|^2.
\end{aligned}$$

Using (35), it follows that

$$\eta\langle x^{k+1}, x^* - \bar{x}\rangle \le \eta\langle x^k, x^* - \bar{x}\rangle + \tfrac{1}{2}(\varphi_k^2 - \varphi_{k+1}^2) + M\zeta_k + \gamma_k,$$

where

$$\gamma_k = \frac{(1+\varrho)R_2}{1-\sigma^2}\lambda_k(\theta(\bar{x}, \varepsilon_k) + \delta_k) + \beta_k^2\left\|g^k\right\|^2$$

is such that $\gamma_k \ge 0$ and $\sum \gamma_k < \infty$ by (5), Proposition 4.1(ii) and (34). Hence

$$a_{k+1} \le a_k + \gamma_k,$$

which proves our claim.

(ii) Assume $\dim H < \infty$. By Proposition 4.2, under the assumption $\sum \lambda_k = \infty$, there exists a subsequence $(x^{k_j})_{j\in\mathbb{N}}$ converging to some $\hat{x}$ in $\operatorname{Argmin}\overline{f}$. By (i) applied to $\bar{x} = \hat{x}$, we conclude that all the sequence $(x^k)_{k\in\mathbb{N}}$ converges to $\hat{x}$.

(iii) By Proposition 4.2, the second condition of Opial's lemma (cf. Lemma 3.1) is fulfilled for $C = \operatorname{Argmin}\overline{f}$, which together with (i) yields the conclusion. □

REMARK 4.3 Theorem 4.1 requires the parameter ε_k to decrease to 0 fast enough in order to have (34), which forces the approximate function $f(\cdot, \varepsilon_k)$ to be close to $\overline{f}$. This is why convergence is ensured towards some point in $\overline{f}$ which may be different to $x^* = \lim_{\varepsilon\to 0} x(\varepsilon)$.

EXAMPLE 4.1 *Penalty/barrier methods in convex programming II* (continued). In the context of Example 3.3, see the references already cited in the Remark 4.2 for results permitting to verify (24), (31) and (32). Concerning (25), the existence of an optimal path $\varepsilon \mapsto x(\varepsilon)$ which is uniformly Lipschitz continuous is classical under strong second-order non-degeneracy conditions using an appropriate application of the Implicit Function Theorem (see [27]). Although such a property is sufficient in order to have (25), it precludes the direct application of this result to problems that admit a multiplicity of optimal solutions. However, the uniform Lipschitz continuity has been established without second-order conditions for some special cases as the log-barrier in linear programming [57, 45]

$$\min_{x\in\mathbb{R}^n} c^t x - \varepsilon\sum_{i=1}^m \log[b_i - a_i^t x],$$

and the exponential penalty in linear programming [23]

$$\min_{x\in\mathbb{R}^n} c^t x + \varepsilon\sum_{i=1}^m \exp\left[(a_i^t x - b_i)/\varepsilon\right].$$

Furthermore, for the former it is possible to take $\theta_k(\bar{x},\varepsilon) = c_0\varepsilon$ where c_0 is a positive constant (see the proof of [20, Corollary 4.1]), while for the latter $\theta_k(\bar{x},\varepsilon) = c_1 \exp[-c_2/\varepsilon]$ where c_1, c_2 are some positive constants (see the proof of [20, Corollary 4.2]). In these cases, to have (34) it suffices that

$$\sum \lambda_k \varepsilon_k < \infty$$

and

$$\forall c > 0, \ \sum \lambda_k \exp[-c/\varepsilon_k] < \infty,$$

respectively. □

4.3 Vanishing strong convexity and slow parametrization Some optimization algorithms consists of closely tracing the optimal path to the optimal solution x^*, using some unconstrained minimization method to obtain a good estimate of $x(\varepsilon_k)$ for a sequence $\varepsilon_k \to 0$ [16, 29, 34, 37, 51]. When ε_k is forced to decrease sufficiently slow to 0, one may expect the iterates of DHP-PPA to get close enough to the optimal path $x(\varepsilon)$ so as to guarantee convergence towards x^*. To obtain a result in this direction, we follow the ideas of [7, 20] by supposing that the family $(f(\cdot,\varepsilon))_{\varepsilon>0}$ satisfies the following local strongly convex condition: for each $\varepsilon > 0$ and for any bounded set K, there exists $\omega_K(\varepsilon) > 0$ such that

$$f(z,\varepsilon) + \langle g, y-z\rangle + \omega_K(\varepsilon)\left\|y-z\right\|^2 \le f(y,\varepsilon), \tag{36}$$

for all $y, z \in K$ and $g \in \partial f(z,\varepsilon)$. Notice that it is allowed that $\omega_K(\varepsilon) \to 0$ as $\varepsilon \to 0$.

THEOREM 4.2 *Suppose that the family $(f(\cdot,\varepsilon))_{\varepsilon>0}$ satisfies* (24), (25), (31), (32) *and* (36). *Let $(x^k)_{k\in\mathbb{N}}$ be generated by* (1)-(4) *applied to $(f(\cdot,\varepsilon_k))_{k\in\mathbb{N}}$, under* (6) *and* (5), *for some sequence $\varepsilon_k \searrow 0$ as $k\to\infty$. If*

$$\sum \lambda_k \omega_K(\varepsilon_k) = \infty, \tag{37}$$

then $\lim_k x^k = x^* = \lim_{\varepsilon\to 0} x(\varepsilon)$.

PROOF. Set $f_k = f(\cdot,\varepsilon_k)$ and $\varphi_k = \left\|x^k - x(\varepsilon_{k-1})\right\|$ for all $k\in\mathbb{N}$. By Proposition 4.1(i), $(\varphi_k)_{k\in\mathbb{N}}$ converges. We claim that its limit is 0. On the one hand, since $(\varphi_k)_{k\in\mathbb{N}}$ converges, the sequence $(x^k)_{k\in\mathbb{N}}$ is bounded. On the other hand, Proposition 4.1 guarantees that $(z^k)_{k\in\mathbb{N}}$ is also bounded. Consequently, there exists a bounded set K such that x^k, z^k and z^k_ρ belong to K for all $k\in\mathbb{N}$. Let ω_K be the modulus of strong convexity associated with K, given by (36). As $0\in\partial f_k(x(\varepsilon_k))$, it follows from (27) that

$$\varphi_{k+1}^2 \le \left\|x^k - x(\varepsilon_k)\right\|^2 - 2\rho_k\beta_k\omega_K(\varepsilon_k)\left\|z^k_\rho - x(\varepsilon_k)\right\|^2 + \tfrac{1}{2}(1-2/\rho_k)\beta_k^2\left\|g^k\right\|^2 + 2\rho_k\beta_k\delta_k.$$

and then

$$\varphi_{k+1}^2 \le \left\|x^k - x(\varepsilon_k)\right\|^2 - R_1(1-\sigma^2)\lambda_k\omega_K(\varepsilon_k)\left\|z^k_\rho - x(\varepsilon_k)\right\|^2 + \mu_k,$$

where μ_k is given by (29). Using the same technique as in the proof of Proposition 4.1 we get

$$\varphi_{k+1}^2 \le \varphi_k^2 + 2M\zeta_k + \zeta_k^2 - R_1(1-\sigma^2)\lambda_k\omega_K(\varepsilon_k)\left\|z^k_\rho - x(\varepsilon_k)\right\|^2 + \mu_k,$$

for ζ_k defined by (26) and $M\ge 0$ such that $\varphi_k \le M$ for all $k\in\mathbb{N}$. Then

$$R_1(1-\sigma^2)\lambda_k\omega_K(\varepsilon_k)\left\|z^k_\rho - x(\varepsilon_k)\right\|^2 \le \varphi_k^2 - \varphi_{k+1}^2 + 2M\zeta_k + \zeta_k^2 + \mu_k.$$

Reasoning exactly as in Proposition 4.1, we can deduce that

$$\sum \lambda_k\omega_K(\varepsilon_k)\left\|z^k_\rho - x(\varepsilon_k)\right\|^2 < \infty.$$

Under the slow parametrization assumption (37), we conclude that $\liminf_k \left\|z^k_\rho - x(\varepsilon_k)\right\| = 0$. Finally, taking $\liminf$ in the following inequality $\left\|x^{k+1} - x(\varepsilon_k)\right\| \le \left\|x^{k+1} - z^k_\rho\right\| + \left\|z^k_\rho - x(\varepsilon_k)\right\|$ allows us to conclude that $\liminf_k \varphi_{k+1} = 0$, which proves the result. □

EXAMPLE 4.2 *Penalty/barrier methods in convex programming III* (continued). We are within the context of Example 3.3 with $H = \mathbb{R}^n$. The functions h_i, $i = 1,\dots,m$ and f are supposed to be twice differentiable as well as ψ in the interior of its domain. Let us assume that for any x, $span\{\nabla h_i(x), i = 1,\dots,m\} = \mathbb{R}^n$. A simple calculation leads to

$$\nabla^2 f(x,\varepsilon) = \nabla^2 f(x) + \sum_{i=1}^m \psi'[h_i(x)/\varepsilon]\nabla^2 h_i(x) + \frac{1}{\varepsilon}\sum_{i=1}^m \psi''[h_i(x)/\varepsilon]\nabla h_i(x)\nabla h_i(x)^t,$$

where $\nabla^2 f$ stands here for the Hessian matrix of f. Let K be a compact set and $x \in K$. Then,

$$\nabla^2 f(x,\varepsilon) \succeq \frac{1}{\varepsilon}\psi''[-\beta_K/\varepsilon]\sum_{i=1}^{m} \nabla h_i(x)\nabla h_i(x)^t,$$

where $-\beta_K := \min_{x\in K,\ i=1..m}\{h_i(x)\}$ so that $\beta_K > 0$ in the worst case. Since $span\{\nabla h_i(x), i = 1..m\} = \mathbb{R}^n$, the matrix $\sum_{i=1}^m \nabla h_i(x)\nabla h_i(x)^t$ is positive definite ; let us denote by $\alpha(x) > 0$ its smallest eigenvalue. We then have

$$\nabla^2 f(x,\varepsilon) \succeq \frac{1}{\varepsilon}\psi''[-\beta_K/\varepsilon]\alpha(x)I.$$

The application $x \mapsto \alpha(x)$ being continuous, there exists $\alpha_K > 0$ such that

$$\nabla^2 f(x,\varepsilon) \succeq \frac{1}{\varepsilon}\psi''[-\beta_K/\varepsilon]\alpha_K I, \quad \forall x \in K. \tag{38}$$

This gives, for instance, $\omega_K(\varepsilon) = \alpha_K \exp[-\beta_K/\varepsilon]$ in the case of the exponential penalty, whereas $\omega_K(\varepsilon) = \alpha_K\varepsilon/\beta_K^2$ for the log-barrier. □

As we have already mentioned (see Remark 4.2), viscosity methods in general do not satisfy the finite length condition in (25). However, under additional conditions, we can prove a convergence result in the finite-dimensional case without assuming the existence of an optimal path. To this end, let us consider $f,\ h:\ H \to \mathbb{R}$ two convex, continuous, bounded from below functions. In addition h is assume to be finite. Set $\overline{S} = \operatorname{Argmin} f$ and $S = \operatorname{Argmin}_{\overline{S}} h$. These two sets are supposed to be nonempty. From now on, we set $f_k = f + \varepsilon_k h$.

Lemma 4.2 *Let* $(x^k)_{k\in\mathbb{N}}$ *be a sequence generated by* (1.1)-(1.4) *applied to the family* $(f_k)_{k\in\mathbb{N}}$ *under the conditions* (5), (6). *Assume that* $\operatorname{Argmin} f \cap \operatorname{Argmin} h \neq \emptyset$. *Then,*

(i) $\sum \beta_k^2 ||g^k||^2 < \infty$.

(ii) If $\inf \lambda_k > 0$ *then* $\lim_k ||g^k|| = 0$ *and* $\lim_k ||z^k - x^k|| = 0$.

(iii) If $\inf \lambda_k > 0$*, if* $\lim_k \delta_k = 0$ *and if the sequence* $(x^k)_{k\in\mathbb{N}}$ *is bounded then any cluster point of* $(x^k)_{k\in\mathbb{N}}$ *belongs to* $\overline{S}$.

Proof. (i) Denote by $\overline{x}$ a point which belongs to $\operatorname{Argmin} f \cap \operatorname{Argmin} h$. We apply Lemma 2.3 to $\overline{x}$. Then, $(2/\rho_k - 1)\beta_k^2||g^k||^2 \le ||x^k - \overline{x}||^2 - ||x^{k+1} - \overline{x}||^2 + 2\beta_k\langle g^k, \overline{x} - z_\rho^k\rangle$. Using now the fact that $g^k/\rho_k \in \partial_{\delta_k} f_k(z_\rho^k)$, we obtain that $(2/\rho_k - 1)\beta_k^2||g^k||^2 \le ||x^k - \overline{x}||^2 - ||x^{k+1} - \overline{x}||^2 + 2\beta_k\rho_k(\delta_k + f_k(\overline{x}) - f_k(z_\rho^k))$. Consequently, $(2/\rho_k - 1)\beta_k^2||g^k||^2 \le ||x^k - \overline{x}||^2 - ||x^{k+1} - \overline{x}||^2 + 2\beta_k\rho_k(\delta_k + f_k(\overline{x}) - \min f_k))$. The assumption $\operatorname{Argmin} f \cap \operatorname{Argmin} h \neq \emptyset$ leads to $f_k(\overline{x}) - \min f_k = 0$ and then $(2/\rho_k - 1)\beta_k^2||g^k||^2 \le ||x^k - \overline{x}||^2 - ||x^{k+1} - \overline{x}||^2 + 2\beta_k\rho_k\delta_k$. Summing over k, we can conclude that: $\sum(2/\rho_k - 1)\beta_k^2||g^k||^2 \le ||x^0 - \overline{x}||^2 + 2\sum \beta_k\rho_k\delta_k < \infty$.
(ii) This is immediate.
(iii) Let us consider a subsequence of $(x^{k_n})_{n\in\mathbb{N}}$ which converges to $\overline{x}$. Then, the sequence $(z_\rho^{k_n})_{n\in\mathbb{N}}$ also converges to $\overline{x}$. We have the following inequality: $f_{k_n}(y) \ge f_{k_n}(z_\rho^{k_n}) + \frac{1}{\rho_{k_n}}\langle g^{k_n}, y - z_\rho^{k_n}\rangle - \delta_{k_n}$, for any $y \in H$. Then, $f(y) + \varepsilon_{k_n}h(y) \ge f(z_\rho^{k_n}) + \varepsilon_{k_n}h(z_\rho^{k_n}) - \frac{1}{\rho_{k_n}}\|g^{k_n}\|\|y - z_\rho^{k_n}\| - \delta_{k_n}$. From (ii), the sequence $(z_\rho^{k_n})_{n\in\mathbb{N}}$ is bounded. Passing to the limit over k, we deduce that: $f(y) \ge f(\overline{x})$, which proves that $\overline{x}$ belongs to $\overline{S}$. □

Theorem 4.3 *Suppose that* $\dim H < \infty$. *Let* $(x^k)_{k\in\mathbb{N}}$ *be a sequence generated by* (1.1)-(1.4) *applied to the family* $(f_k)_{k\in\mathbb{N}}$ *under the conditions* (5), (6). *Assume that:*

(i) $\operatorname{Argmin} f \cap \operatorname{Argmin} h \neq \emptyset$.

(ii) $\inf \lambda_k > 0$.

(iii) $\sum \lambda_k\varepsilon_k = \infty$.

Then, $(x^k)_{k\in\mathbb{N}}$ *converges to a point of* S.

Proof. We will adapt some arguments of [18].
Let $\overline{x}$ be in S and set $\varphi_k = \frac{1}{2}||x^k - \overline{x}||^2$. We are going to prove that the sequence $(||x^k - \overline{x}||)_{k\in\mathbb{N}}$

converges. Then, we have $\varphi_{k+1} - \varphi_k \leq \beta_k\rho_k\delta_k + \beta_k\rho_k(f_k(\overline{x}) - f_k(z^k_\rho))$, and since $\overline{x} \in S$, $\varphi_{k+1} - \varphi_k \leq \beta_k\rho_k\delta_k + \beta_k\rho_k[\min f - f(z^k_\rho) + \varepsilon_k(\min_{\overline{S}} h - h(z^k_\rho))]$. Clearly,

$$\varphi_{k+1} - \varphi_k \leq \beta_k\rho_k\delta_k + \beta_k\rho_k\varepsilon_k(\min_{\overline{S}} h - \min h).$$

However, it is easy to prove that $\operatorname{Argmin} f \cap \operatorname{Argmin} h \neq \emptyset$ leads to $\min_{\overline{S}} h = \min h$. Thus $\varphi_{k+1} - \varphi_k \leq \beta_k\rho_k\delta_k$, and the sequence $(\varphi_k)_{k\in\mathbb{N}}$ converges.
Let us now prove that $d(x^k, S)$ converges to 0. Set $d_k = \frac{1}{2}d(x^k, S)^2$. Then, $d_k = \frac{1}{2}||x^k - p_S(x^k)||^2$, where p_S denotes the projection operator on S. Consequently, $d_k = \frac{1}{2}||x^k - p_S(x^k) - (x^{k+1} - p_S(x^{k+1}))||^2 + \frac{1}{2}||x^{k+1} - p_S(x^{k+1})||^2 + \langle x^{k+1} - p_S(x^{k+1}), x^k - p_S(x^k) + x^{k+1} - p_S(x^{k+1})\rangle$. Using the projection property, we obtain $d_k \geq d_{k+1} + \langle x^k - x^{k+1}, x^{k+1} - p_S(x^{k+1})\rangle$. We have $\langle x^{k+1} - x^k, x^{k+1} - p_S(x^{k+1})\rangle = \beta_k\langle g^k, p_S(x^{k+1}) - z^k_\rho\rangle + \beta_k\langle g^k, z^k_\rho - x^{k+1}\rangle$. On the one hand $\langle g^k, p_S(x^{k+1}) - z^k_\rho\rangle \leq \rho_k(f_k(p_S(x^{k+1})) - \phi_k(z^k_\rho) + \delta_k)$ and then $\langle g^k, p_S(x^{k+1}) - z^k_\rho\rangle \leq \rho_k\varepsilon_k(\min_{\overline{S}} h - h(z^k_\rho)) + \delta_k\rho_k$. On the other hand, it is easy to see that $\langle g^k, z^k_\rho - x^{k+1}\rangle = \beta_k||g^k||^2(1 - 1/\rho_k)$. Finally,

$$d_{k+1} - d_k \leq \beta_k\rho_k\varepsilon_k(\min_{\overline{S}} h - h(z^k_\rho)) + \beta_k\delta_k\rho_k + \beta_k^2||g^k||^2(1 - 1/\rho_k) \tag{39}$$

Let us distinguish two cases:
Case1: $\exists n_0 \in \mathbb{N}\ \forall n \geq n_0\ h(z^k_\rho) > \min_{\overline{S}} h$. Then (39) becomes $d_{k+1} - d_k \leq \beta_k\delta_k\rho_k + \beta_k^2||g^k||^2(1 - 1/\rho_k)$. Note that the second member of the inequality has a finite sum by Lemma 4.2. Consequently, $(d_k)_{k\in\mathbb{N}}$ converges. Let us now prove that it converges to 0. Inequality (39) leads now to $\beta_k\rho_k\varepsilon_k(h(z^k_\rho) - \min_{\overline{S}} h) \leq d_k - d_{k+1} + \beta_k\delta_k\rho_k + \beta_k^2||g^k||^2(1 - 1/\rho_k)$ and then, summing over k, $\sum \beta_k\rho_k\varepsilon_k(h(z^k_\rho) - \min_{\overline{S}} h) < \infty$. Under the assumption of slow convergence we then obtain that $\liminf_k h(z^k_\rho) = \min_{\overline{S}} h$. Using Lemma 4.2 (ii), we obtain that $\liminf_k h(x^k) = \min_{\overline{S}} h$. Let us consider a sequence $(x^{k_n})_{n\in\mathbb{N}}$ which converges to $\liminf_k h(x^k)$. Since it is a bounded sequence, we can extract a converging subsequence $(x^{k_{n_p}})_{p\in\mathbb{N}}$ of $(x^{k_n})_{n\in\mathbb{N}}$, whose limit is denoted by $\overline{x}$. Since h is continuous, we get $\lim_{p\to\infty} h(x^{k_{n_p}}) = h(\overline{x})$. Applying Lemma 4.2 (iii), we finally obtain that $\overline{x}$ belongs to $\overline{S}$. Finally, $\overline{x}$ belongs to S, which allows to conclude that $\lim_k d_k = 0$.
Case2: $\forall n_0 \in \mathbb{N}\ \exists n \geq n_0\ h(z^k_\rho) \leq \min_{\overline{S}} h$. Let us denote $\tau_n = \max\{k \mid k \leq n \text{ and } h(z^k_\rho) \leq \min_{\overline{S}} h)\}$. Note that $\lim_n \tau_n = \infty$. We easily get

$$d_n \leq d_{\tau_n} + \sum_{k=\tau_n}^{\infty} \lambda_k\delta_k + \beta_k^2||g^k||^2(1 - 1/\rho_k). \tag{40}$$

Let us now prove that the sequence $(d_{\tau_n})_{n\in\mathbb{N}}$ converges to 0. Let us consider a subsequence $(x^{k_n})_{n\in\mathbb{N}}$ of $(x^n)_{n\in\mathbb{N}}$ which converges to $\overline{x}$. Then, the sequence $(z^{k_n}_\rho)_{n\in\mathbb{N}}$ also converges to $\overline{x}$ (Lemma 4.2 (ii)). Since, $z^{k_n}_\rho \in [h \leq \min_{\overline{S}} h]$, $\overline{x}$ eventually belongs to $[h \leq \min_{\overline{S}} h]$. From Lemma 4.2 (iii), we know that $\overline{x}$ belongs to $\overline{S}$ and it is straighforward to conclude that $\overline{x}$ belongs to S. All the cluster points of $(d_{\tau_n})_{n\in\mathbb{N}}$ are then equal to 0. Finally, we apply (40) to conclude that $(d_n)_{n\in\mathbb{N}}$ converges to 0.
Since $(x^k)_{k\in\mathbb{N}}$ is bounded, we can extract a subsequence from $(x^k)_{k\in\mathbb{N}}$ which converges. As $(d_k)_{k\in\mathbb{N}}$ converges to 0, the limit point $\overline{x}$ of this subsequence belongs to S. From (i'), the whole sequence $(||x^k - \overline{x}||)_{k\in\mathbb{N}}$ converges and the conclusion follows. □

Remark 4.4 It would be interesting to obtain a slow parametrization result for viscosity methods without the very stringent condition $\operatorname{Argmin} f \cap \operatorname{Argmin} h \neq \emptyset$.

5. Extension: finding zeros of maximal monotone operators In this section we briefly discuss versions of some results of this paper for a sequence of maximal monotone[2] operators $(A_k)_{k\in\mathbb{N}}$ converging in a certain sense to a maximal monotone operator $\overline{A}$. Let us first write the corresponding proximal step: given $x^k \in H$, $\lambda_k > 0$ and $\rho_k \in (0, 2)$, find $z^k \in H$ such that

$$(z^k - x^k)/\lambda_k + g^k = \xi^k, \text{ for } g^k \in \rho_k A_k(z^k_\rho), \tag{41}$$

where the residue ξ^k satisfies (3). The projection step remains exactly the same: if $g^k = 0$ then set $x^{k+1} = x^k$; otherwise, take

$$x^{k+1} = x^k - \beta_k g^k, \text{ with } \beta_k = \langle g^k, x^k - z^k\rangle / \left\|g^k\right\|^2. \tag{42}$$

[2]Recall that a set-valued map $A : H \rightrightarrows H$ is a *maximal monotone operator* if A is monotone, i.e., $\forall x, y \in H$, $\forall v \in A(x)$, $\forall w \in A(y)$, $\langle v - w, x - y\rangle \geq 0$, and if the graph $GrA = \{(x, v) \in H \times H \mid v \in A(x)\}$ is not properly contained in the graph of any other monotone operator.

Let us now consider a family of approximated problems:

$$(P^\varepsilon) \qquad\qquad 0 \in A(x, \varepsilon),$$

where each operator $A(., \varepsilon)$ is maximal monotone. Set $S_\varepsilon = A(., \varepsilon)^{-1}(0)$. The following conditions will help us to prove the convergence of the algorithm:

$$\text{There exists a function } x : (0, \varepsilon_0] \to H \text{ such that } \forall \varepsilon \in (0, \varepsilon_0],\ x(\varepsilon) \in S_\varepsilon. \tag{43}$$

$$\text{The optimal path } x(\cdot) \text{ is absolutely continuous on } (0, \varepsilon_0] \text{ and } \int_0^{\varepsilon_0} \left| \frac{dx}{d\varepsilon} \right| d\varepsilon < \infty. \tag{44}$$

$$\text{There exists a maximal operator } \overline{A} \text{ such that } \lim_{\varepsilon \to 0} x(\varepsilon) = x^\star, \text{ where } 0 \in \overline{A}(x^\star). \tag{45}$$

Remark 5.1 Let $(x_k)_{k\in\mathbb{N}}$ be the sequence generated by (41)-(42) applied to $A_k = A(., \varepsilon_k)$ for some sequence $\varepsilon_k \to 0$. On the one hand, from Lemma 2.2, $\langle g^k, x^k - z^k_\rho \rangle = \frac{1}{\rho_k} \langle g^k, x^k - z^k \rangle > 0$. On the other hand, the maximal monotonicity of $A(., \varepsilon_k)$ leads to $\langle g^k, x(\varepsilon_k) - z^k_\rho \rangle \leq 0$. Then, the points x^k and $x(\varepsilon_k)$ are strictly separated by the hyperplane $P_k := \{x \in H \mid \langle g^k, x - z^k_\rho \rangle = 0\}$.

We have the following preliminary result, whose proof is similar to the proof of Proposition 4.1:

Proposition 5.1 *Assume that* (43) *and* (44) *hold. Given a sequence* $(\varepsilon_k)_{k\in\mathbb{N}}$, *let* $(x_k)_{k\in\mathbb{N}}$ *be the sequence generated by* (41)-(42), *applied to* $A_k = A(., \varepsilon_k)$ *under the conditions* (5) *and* (6). *Then the sequence* $(\|x^{k+1} - x(\varepsilon_k)\|)_{k\in\mathbb{N}}$ *converges,* $\lim_k \lambda_k ||g^k|| = 0$, *and* $\lim_k ||z^k_\rho - x^k|| = 0$.

Let us recall that a sequence $(A_k)_{k\in\mathbb{N}}$ graph-converges to A if for any $(x, y) \in Graph(A)$ there exists $(x_k, y_k) \in Graph(A_k)$ with $x_k \to x$ and $y_k \to y$. For more details on graph-convergence, we refer the reader to [6]. We can distinguish two kind of parametrization, a fast one and a slow one. Let us first consider the fast parametrization case:

Theorem 5.1 *Assume that* (43)-(45) *hold. Let* $\lambda_k \geq 0$ *be such that* $\inf \lambda_k > 0$. *Given a sequence* $(\varepsilon_k)_{k\in\mathbb{N}}$, *let* $(x_k)_{k\in\mathbb{N}}$ *be the sequence generated by* (41)-(42) *applied to* $A_k = A(., \varepsilon_k)$, *under the conditions* (5) *and* (6). *We suppose that* $(A(., \varepsilon_k))_{k\in\mathbb{N}}$ *graph-converges to* $\overline{A}$. *Besides, we assume that for each* $\bar{x}$ *of* $\overline{S}$, *there exist* $\eta > 0$ *and* ζ_k *such that*

$$\zeta_k \in A(x(\varepsilon_k) + \eta(\bar{x} - x^*), \varepsilon_k), \tag{46}$$

Under the fast parametrization condition

$$\sum \lambda_k \|\zeta_k\| < \infty,$$

the sequence $(x^k)_{k\in\mathbb{N}}$ *weakly converges to a point of* $\overline{A}^{-1}(0)$.

Proof. We are going to apply Opial's lemma (3.1). Note first that the first condition is obtained as in the proof of Theorem 4.1. Since Proposition 4.2 is a key-point on proving that the second condition of Opial's lemma is fulfilled, the adaptation to maximal monotone case is then not straighforward and we resort to graph-convergence to conclude the proof: let $(x^{k_j})_{j\in\mathbb{N}}$ be a sequence which weakly converges to a point x^∞. Then $(z^{k_j}_\rho)_{j\in\mathbb{N}}$ weakly converges to x^∞ (Theorem 5.1). We also know that $(g^{k_j})_{j\in\mathbb{N}}$ converges to 0 (by Theorem 5.1 and $\inf \lambda_k > 0$) and $g^{k_j} \in A(z^{k_j}_\rho, \varepsilon_{k_j})$. Using Proposition 3.59 of [6] related to graph-convergence, we can conclude that $0 \in A(x^\infty)$, this proves the second condition of Opial's lemma. □

If the parametrization of the family is slow, the algorithm converges to the limit point of $(x(\varepsilon_k))_{k\in\mathbb{N}}$:

Theorem 5.2 *Assume that* (43), (44) *and* (45) *hold. Consider a sequence* $(\varepsilon_k)_{k\in\mathbb{N}}$ *and the sequence* $(x^k)_{k\in\mathbb{N}}$ *generated by* (41)-(42), *applied to* $A_k = A(., \varepsilon_k)$. *We also assume that the estimates* (5) *and* (6) *are true. We suppose that the family* $A(., \varepsilon_k)$ *satisfies a strong monotonicity condition with rate* $w(\varepsilon_k)$ *that is*

$$\langle u - v, x - y \rangle - \omega(\varepsilon_k) \|x - y\|^2 \geq 0, \tag{47}$$

for any $u \in A(x, \varepsilon_k)$*,* $v \in A(y, \varepsilon_k)$*, under the slow parametrization condition:*

$$\sum \lambda_k \omega(\varepsilon_k) = \infty. \tag{48}$$

Then the sequence $(x^k)_{k\in\mathbb{N}}$ *converges to* x^**.*

PROOF. Combining Lemma 2.3 applied to $x(\varepsilon_k)$ with the assumption of strong monotonicity (47), we obtain the following inequality: $\varphi_{k+1}^2 \leq \left\|x^k - x(\varepsilon_k)\right\|^2 - 2\rho_k\beta_k\omega(\varepsilon_k)\left\|z_\rho^k - x(\varepsilon_k)\right\|^2 + \frac{1}{2}(1-2/\rho_k)\beta_k^2\left\|g^k\right\|^2$. The rest of the proof is similar to the proof of Theorem 4.2 with $\mu_k = 0$. □

Acknowledgments. This research was partially supported by FONDECYT Grant 1050706, FONDAP in Applied Mathematics and Millennium Scientific Initiative. The third author was also supported by ECOS/CONICYT.

References

[1] Alart P., Lemaire B., *Penalization in nonclassical convex programming via variational convergence*, Math. Program. 51 (1991), pp. 307-331.

[2] Alvarez F., *Absolute minimizer in convex programming by exponential penalty*, J. Convex Anal. 7 (2000), pp. 197-202.

[3] Alvarez F., *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space*, SIAM J. Optim. 14 (2004), pp. 773-782.

[4] Alvarez F., Cominetti R., *Primal and dual convergence of a proximal point exponential penalty method for linear programming*, Math. Program. 93 (2002), no. 1, Ser. A, pp. 87-96.

[5] Attouch H., *Viscosity solutions of minimization problems*, SIAM J. Optimization 6 (1996), pp. 769-806.

[6] Attouch H. "Variational convergence for functions and operators". Pitman, Boston, 1984.

[7] Attouch H., Cominetti R., *A dynamical approach to convex minimization coupling approximation with the steepest descent method*, J. Diff. Equations 128 (1996), pp. 519-540.

[8] Attouch H., Cominetti R., *L^p approximation of variational problems in L^1 and L^∞*, Nonlinear Anal. 36 (1999), pp. 373-399.

[9] Auslender A., *Numerical methods for nondifferentiable convex optimization*, Math. Program. Studies 30 (1987), pp. 102-127.

[10] Auslender A., Cominetti R., Haddou M., *Asymptotic analysis for penalty methods in convex and linear programming*, Math. Oper. Res. 22 (1997), pp. 43-62.

[11] Auslender A., Crouzeix J.P., Fedit P., *Penalty-proximal methods in convex programming*, J. Optim. Theory Appl. 55 (1987), pp. 1-21.

[12] Bahraoui M.A., Lemaire B., *Convergence of diagonally stationary sequences in convex optimization*, Set-Valued Anal. 2 (1994), pp. 49-61.

[13] Bakushinskii A.B., Goncharskii A.V., "Iterative Methods for Ill-posed Problems", Kluwer Academic Publishers, Dordrecht, 1994.

[14] Bertsekas D.P., *Approximation procedures based on the method of multipliers*, J. Optim. Theory Appl. 23 (1977), pp. 487-510.

[15] Bertsekas D.P., "Constrained Optimization and Lagrange Multiplier Methods", Academic Press, New York, 1982.

[16] Bonnans J.F., Gilbert J., Lemaréchal C., Sagastizábal, C., "Numerical optimization. Theoretical and practical aspects". Translated and revised from the 1997 French original. Universitext. Springer-Verlag, Berlin, 2003.

[17] Brézis H., "Opérateurs maximaux monotones", Mathematics Studies, 5, North Holland, 1973.

[18] Cabot A., *Proximal point and steepest descent algorithm controlled by slowly vanishimg term. Applications to hierarchical minimization*, prepublication of the University of Limoges, France, 06 (2003).

[19] Champion T., *Tubularity and asymptotic convergence of penalty trajectories in convex programming*, SIAM J. Optim. 13 (2002), pp. 212-227.

[20] Cominetti R., *Coupling the proximal point algorithm with approximation methods*, J. Optim. Theory Appl. 95 (1997), pp. 581-600.

[21] Cominetti R., *Nonlinear averages and convergence of penalty trajectories in convex programming*, in Ill-Posed Variational Problems and Regularization Techniques (Trier, 1998), Lecture Notes in Econom. and Math. Systems 477, Springer, Berlin, 1999, pp. 65-78.

[22] Cominetti R., Courdourier M., *Coupling general penalty schemes for convex programming with the steepest descent and the proximal point algorithm*, SIAM J. Optim. 13 (2002), pp. 745-765.

[23] Cominetti R., San Martín J., *Trajectory analysis for the exponential penalty method in linear programming*, Math. Program. 67 (1994), pp. 169-187.

[24] Correa R., Lemarechal C., *Convergence of some algorithms for convex minimization*, Math. Program. 62 (1993), pp. 261-275.

[25] Eckstein J., Bertsekas D.P., *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program. 55 (1992), pp. 293-318.

[26] Eckstein J., Ferris M.C., *Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control*, INFORMS J. Comput. 10 (1998), pp. 218-235.

[27] Fiacco A., McCormick G., "Nonlinear Programming. Sequential unconstrained minimization techniques". Revised second edition of the 1968 original. Classics in Applied Mathematics, SIAM, 1990.

[28] Facchinei, F., Pang, J.-S., "Finite dimensional Variational Inequalities and Complementarity Problems", Springer-Verlag, Berlin, 2003.

[29] Forsgren A., Gill P.E., Wright M.H., *Interior methods for nonlinear optimization*, SIAM Review 44 (2002), pp. 525-597.

[30] Gárciga O.R., Iusem A., Svaiter B.F., *On the need for hybrid steps in hybrid proximal point methods*, Oper. Res. Letters 29 (2001), pp. 217-220.

[31] Gol'shtein E.G., *A method for the modification of monotone mapings*, Ekon. i. Matem. Metody (Matecon) 11 (1975), pp. 1144-1159.

[32] Gol'shtein E.G., Tret'yakov N.V., *Modified Lagrangians in convex programming and their generalizations*, Math. Program. Study, 10 (1979), pp. 86-97.

[33] Gol'shtein E.G., Tret'yakov N.V., "Modified Lagrangians and Monotone Maps in Optimization", John Wiley and Sons, New York, 1996.

[34] Gonzaga C.G., *Path-following methods for linear programming*, SIAM Review 34 (1992), pp. 167-224.

[35] Güler O., *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optim. 29 (1991), pp. 403-419.

[36] Hiriart-Urruty J.-B., Lemaréchal C., "Convex analysis and minimization algorithms. II. Advanced theory and bundle methods". Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 306. Springer-Verlag, Berlin, 1993.

[37] Jarre F., Saunders M.A., *A practical interior-point method for convex programming*, SIAM J. Optim. 5 (1995), no. 1, 149–171.

[38] Kaplan A., *On a convex programming method with internal regularization*, Soviet Mathematics Doklady, Vol. 19 (1978), pp. 795-799.

[39] Kaplan A., Tichatschke R., *Proximal methods in view of interior-point strategies*, J. Optim. Theory Appl. 98 (1998), pp. 399-429.

[40] Konnov, I.V., *Combined relaxation methods for finding equilibrium points and solving related problems*, Russian Math. (Iz. VUZ), 37 (1993), pp. 44-51.

[41] Konnov, I.V., "Combined Relaxation Methods for Variational Inequalities", Springer-Verlag, Berlin, 2001.

[42] Lemaire B., *Coupling optimization methods and variational convergence*, Trends in Mathematical Optimization, International Series of Numerical Mathematics, Birkhäuser, Bassel (1988), pp. 163-179.

[43] Lemaire B., *About the convergence of the proximal method*, Advances in Optimization, Proc. Lambrecht 1991, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, New York (1992), pp. 39-51.

[44] Martinet B., *Régularisation d'inéquations variationnelles par approximations successives*, Revue Française d'Informatique et Recherche Operationnelle, 4 (1970), pp. 154-159.

[45] Megiddo N., *Pathways to the optimal set in linear programming*. Progress in mathematical programming (Pacific Grove, CA, 1987), pp. 131-158, Springer, New York, 1989.

[46] Monteiro R., Zhou F., *On the existence and convergence of the central path for convex programming and some duality results*, Comput. Optim. Appl. 10 (1998), pp. 51-77.

[47] Mouallif K., *Sur la convergence d'une méthode associant pénalisation et réguralisation*, Bulletin de la Societé Royale des Sciences de Liège, Vol. 56 (1987), pp. 175-180.

[48] Mouallif K., Tossings P., *Une méthode de pénalisation exponentielle associée à une régularisation proximale*, Bulletin de la Societé Royale des Sciences de Liège, Vol. 56 (1987), pp. 181-190.

[49] Mouallif K., Tossings P., *Variational metric and exponential penalization*, J. Optim. Theory Appl. 67 (1990), pp. 185-192.

[50] Moudafi A., *Coupling proximal algorithm and Tikhonov method*, Nonlinear Times and Digest, Vol. 1 (1994), pp. 203-210.

[51] Nesterov Y., Nemirovskii A., *Interior-point polynomial algorithms in convex programming*. SIAM Studies in Applied Mathematics, 13. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[52] Opial Z., *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. of the American Math. Society 73, (1967), pp. 591-597.

[53] Rockafellar R.T., *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877-898.

[54] Rockafellar R. T., *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res. 1 (1976), pp. 97-116.

[55] Solodov M.V., Svaiter B.F., *A hybrid projection-proximal point algorithm*, J. Convex Anal. 6 (1999), pp. 59-70.

[56] Solodov M.V., Svaiter B.F., *A unified framework for some inexact proximal point algorithms*, Num. Funct. An. Opt. 22 (2001), pp. 1013-1035.

[57] Sonnevend G., *An "analytical centre" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming*, System modelling and optimization (Budapest, 1985), pp. 866-875, Lecture Notes in Control and Inform. Sci., 84, Springer, Berlin, 1986.

[58] Tikhonov A., Arsenine V., "Méthodes de résolution de problèmes mal posés", French translation of the 1974 Russian edition, Mir, Moscow, 1976.