Math. Program., Ser. A 93: 87–96 (2002)

Digital Object Identifier (DOI) 10.1007/s10107-002-0295-0

F. Alvarez · R. Cominetti

Primal and dual convergence of a proximal point exponential penalty method for linear programming

Received: May 2000 / Accepted: November 2001
Published online June 25, 2002 –

Abstract. We consider the diagonal inexact proximal point iteration

$$\frac{u^k - u^{k-1}}{\lambda_k} \in -\partial_{\varepsilon_k} f\big(u^k, r_k\big) + \nu^k$$

where $f(x, r) = c^{\mathrm{T}}x + r\sum \exp[(A_i x - b_i)/r]$ is the exponential penalty approximation of the linear program $\min\{c^{\mathrm{T}}x : Ax \leq b\}$. We prove that under an appropriate choice of the sequences λ_k, ε_k and with some control on the residual ν^k, for every $r_k \to 0^+$ the sequence u^k converges towards an optimal point u^∞ of the linear program. We also study the convergence of the associated dual sequence $\mu_i^k = \exp[(A_i u^k - b_i)/r_k]$ towards a dual optimal solution.

Key words. proximal point – exponential penalty – linear programming

1. Introduction and main results

Consider a linear program of the form

$$\min_{x\in\mathbb{R}^n} \{c^{\mathrm{T}}x : Ax \leq b\} \tag{P}$$

with $c \in \mathbb{R}^n$, A an $m \times n$ matrix, and $b \in \mathbb{R}^m$, such that (P) has a nonempty and bounded optimal solution set $S(P)$. The exponential penalty approximates (P) by the family of smooth unconstrained problems

$$\min_{x\in\mathbb{R}^n} \left\{ f(x, r) := c^{\mathrm{T}}x + r \textstyle\sum_{i=1}^m \exp[(A_i x - b_i)/r] \right\} \tag{P_r}$$

where A_i denote the rows of A, and $r > 0$ is a *penalty* parameter. For each $r > 0$ the function $f(\cdot, r)$ is strictly convex and coercive, so that (P_r) has a unique optimal solution $x(r)$ which converges when $r \to 0$ towards the *centroid* of the optimal face of the linear program, denoted $x^* \in S(P)$. We refer to [8] for a detailed study of the path $r \mapsto x(r)$, and to [2] for extensions to nonlinear programming.

The standard approach in penalty-based methods consists in closely tracing the optimal path $r \mapsto x(r)$ towards the optimal solution x^* using some unconstrained

F. Alvarez, R. Cominetti: Universidad de Chile, Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático, Casilla 170/3, Correo 3, Santiago, Chile

Partially supported by Conicyt-Chile under Fondecyt grants 1990884 and 1981052 respectively. The authors also wish to thank the french-chilean scientific cooperation program ECOS, as well as the Laboratoire d'Econométrie de L'Ecole Polytechnique where part of this research was carried out.

Mathematics Subject Classification (2000): 90C25

minimization technique to estimate $x(r_k)$ for a sequence of parameters $r_k \to 0$. However, approximating $x(r_k)$ with precision may be expensive and to some extent pointless if the goal is simply to find *any* optimal solution of (P). It is then natural to ask whether an accurate approximation of $x(r_k)$ is really necessary for generating a method converging to an optimal solution of (P). A possible framework for stating the question more precisely is to consider algorithms that perform a prescribed number of iterations of a descent method applied to $f(\cdot, r_k)$, and then update the parameter r_k to r_{k+1}. Global convergence of such economical algorithms is nontrivial because one may no longer expect the iterates to remain close to the optimal path $x(r_k)$ so as to guarantee convergence towards x^*, unless the sequence r_k is forced to decrease slowly towards 0 affecting eventually the overall efficiency of the method.

The aim of this paper is to establish global convergence of a particular iterative method for solving (P) in which the exponential penalty scheme is coupled with the proximal point algorithm [10, 16, 21]. More precisely, we consider a method performing *a single* inexact prox iteration for each value of the penalty parameter r_k, that is to say, we study the sequences u^k satisfying[1]

$$\frac{u^k - u^{k-1}}{\lambda_k} \in -\partial_{\varepsilon_k} f\big(u^k, r_k\big) + \nu^k \qquad (PE)$$

with u^0 a given starting point, $r_k > 0$ a sequence of penalty parameters tending to 0, $\lambda_k > 0$ a sequence of stepsizes, $\varepsilon_k \geq 0$ a tolerance for the computation of approximate subgradients, and $\nu^k \in \mathbb{R}^n$ the residual in the inexact resolution of the prox iteration. We remark that (PE) is a *generic* numerical method since we do not specify a procedure for computing the next iterate u^k. In particular one may compute u^k by using the Bundle method (with $\varepsilon_k > 0$ and $\nu^k = 0$) or by applying some iterations of a standard descent method for the unconstrained minimization of the smooth strongly convex function $f_k(x) = f(x, r_k) + \frac{1}{2\lambda_k}\|x - u^{k-1}\|^2$ (with $\varepsilon_k = 0$ and $\nu^k = \nabla f_k(u^k)$). Our main result on the convergence of (PE) is as follows.

Theorem 1. *Suppose that the optimal set $S(P)$ of the linear program (P) is nonempty and bounded and let u^k be a sequence satisfying (PE) with*

(a) $r_k \to 0$,

(b) $\sum \lambda_k = \infty$,

(c) $\sum[\varepsilon_k + \|\nu^k\|]\lambda_k < \infty$

Then u^k converges when $k \to \infty$ towards some $u^\infty \in S(P)$.

Similar parametric penalty-proximal methods in the context of linear and nonlinear convex programming have been considered by several authors. Auslender et al. studied in [4] the combination of the proximal point algorithm with some exterior penalty methods as the classical quadratic penalty, through an iterative scheme analogous to (PE). This idea had previously been considered by Kaplan [12] but for a class of interior penalty methods (see [13] for more recent results). Following this approach, Lemaire [14] and Alart and Lemaire [1] extended the convergence results

[1] $\partial_\varepsilon f(x, r)$ denotes the ε-subdifferential of the convex function $f(\cdot, r)$.

of [4] via variational convergence methods and considered other penalty functions (see also [5]). We may also mention the coupling of the proximal iteration with other regularization-penalty methods [17] and with Tikhonov regularization [20]. It is worth pointing out that Mouallif and Tossings [18, 19] proved a convergence result for a two-parameter exponential penalty proximal algorithm for convex programs. See [15] for other general results concerning inexact versions of the proximal iteration similar to (PE), and [7] for general results for diagonal prox methods and their application to Tikhonov regularization as well as to the log-barrier and the exponential penalty in linear programming.

A common point to all the works mentioned above is that additional restrictive conditions are imposed on the penalty sequence $\{r_k\}$ in order to attain convergence. The key point in Theorem 1 is precisely that convergence is ensured for *every* sequence $r_k \to 0$. In particular, this provides a theoretical support for algorithms in which the penalty parameter is updated *on-line* (a strategy to avoid numerical instabilities), and for which the sequence r_k is not known a priori so it may be difficult or even impossible to ensure that it satisfies any extra condition.

Let us turn our attention to the dual problem

$$\min_{\lambda \in \mathbb{R}^m} \{b^{\mathrm{T}}\lambda : c + A^{\mathrm{T}}\lambda = 0, \lambda \geq 0\}. \tag{D}$$

Since $S(P)$ is nonempty and bounded, this dual is strictly feasible (there exists $\lambda^0 > 0$ such that $c + A^{\mathrm{T}}\lambda^0 = 0$) and its optimal set $S(D)$ is nonempty. Observe that the optimality condition for (P_r) can be written as

$$0 = \nabla f(x(r), r) = c + \sum_{i=1}^{m} \lambda_i(r) A_i = c + A^{\mathrm{T}}\lambda(r),$$

with the multiplier vector $\lambda(r) = (\lambda_i(r))_{i=1}^m$ given by $\lambda_i(r) = \exp[(A_i x(r) - b_i)/r]$. It turns out that this $\lambda(r)$ is the unique optimal solution of the following barrier approximation of the dual problem

$$\min_{\lambda \in \mathbb{R}^m} \left\{ b^{\mathrm{T}}\lambda + r \sum_{i=1}^{m} \lambda_i (\ln \lambda_i - 1) : A^{\mathrm{T}}\lambda + c = 0, \lambda \geq 0 \right\}, \tag{D_r}$$

and this $\lambda(r)$ converges when $r \to 0$ towards a particular dual solution $\lambda^* \in S(D)$ (see [8] for details). Similarly, we may associate with u^k a multiplier sequence μ^k by

$$\mu_i^k = \exp\left[\left(A_i u^k - b_i\right)/r_k\right], \; i = 1, ..., m. \tag{1}$$

The question is whether μ^k converges towards a dual optimal solution $\mu^\infty \in S(D)$. It turns out that under mild assumptions on the sequence r_k this convergence holds with $\mu^\infty = \lambda^*$. Namely, in Sect. 3 we prove the following.

Theorem 2. *In addition to the assumptions of Theorem* 1*, suppose that:*

(d) r_k *is nonincreasing with* $q_k = (r_{k-1} - r_k)/(r_{k-1}\lambda_k)$ *bounded,*

(e) $\nu^k \to 0$

(f) $\varepsilon_k/r_k \to 0$,

(g) *either* $\varepsilon_k/\lambda_k \to 0$ *or* $\sum \varepsilon_k/r_k < \infty$.

Then, the sequence μ^k *given by* (1) *converges towards* $\lambda^* \in S(D)$ *when* $k \to \infty$.

This type of result ensuring dual convergence for a purely primal penalty method is rather unfrequent in the literature. We stress the fact that neither primal nor dual uniqueness for the linear program is assumed, so that no strong second order sufficient conditions are present. As far as we know Theorem 2 is the first dual convergence result in such a general framework.

Theorems 1 and 2 (and partly their proofs) are inspired in a recent work [6] which studies the asymptotic behavior of the flow generated by the differential equation

$$\frac{du}{dt} = -\nabla f(u(t), r(t)),$$

where $r(t) > 0$ is a given function tending to 0 as $t \to \infty$. This equation is a continuous version of the steepest descent method coupled with the exponential penalty, and (PE) corresponds to an implicit discretization scheme for its numerical integration.

2. Proof of Theorem 1: primal convergence

2.1. A general distance-convergence result

In this section we prove a general result on the asymptotic behavior of sequences generated by diagonal proximal-type iterative schemes. Let E be a *finite dimensional* Euclidean space with scalar product and norm denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ respectively. As usual, $\Gamma_0(E)$ denotes the class of all closed proper convex functions defined on E.

Theorem 3. *Let* $\varphi_k \in \Gamma_0(E)$ *and* $\alpha_k \geq 0$. *Let* $w^k \in E$ *be a sequence satisfying*

$$\frac{w^k - w^{k-1}}{\lambda_k} \in -\partial_{\varepsilon_k}[\alpha_k\varphi_k](w^k) + \nu^k \tag{2}$$

with $\lambda_k > 0$, $\varepsilon_k \geq 0$ *and* $\nu^k \in E$ *such that* $\sum \alpha_k\lambda_k = \infty$ *and* $\sum[\varepsilon_k + \|\nu^k\|]\lambda_k < \infty$. *Suppose there exists* $\varphi_\infty \in \Gamma_0(E)$ *with*

(H1) $v_\infty := \inf_y \varphi_\infty(y)$ *finite and* $S_\infty := \operatorname{Argmin}(\varphi_\infty)$ *nonempty and bounded.*

(H2) $\varphi_\infty(y_\infty) \leq \liminf_j \varphi_{k_j}(y_j)$ *for all sequences* $k_j \to \infty$ *and* $y_j \to y_\infty$.

(H3) $v_k := \sup\{\varphi_k(y) : y \in S_\infty\} \to v_\infty$ *as* $k \to \infty$.

Then $\operatorname{dist}(w^k, S_\infty)$ *tends to 0 as* $k \to \infty$.

Proof. Let $\theta_k = \frac{1}{2}\|w^k - y^k\|^2$ with y^k the projection of w^k onto S_∞. We must prove that $\theta_k \to 0$. To this end we note that $\frac{1}{2}\|u\|^2 - \frac{1}{2}\|v\|^2 \leq \langle u - v, u\rangle$ so that (2) implies

$$\begin{aligned}\theta_k - \theta_{k-1} &\leq \tfrac{1}{2}\|w^k - y^{k-1}\|^2 - \tfrac{1}{2}\|w^{k-1} - y^{k-1}\|^2\\ &\leq \lambda_k\left\langle \tfrac{w^k - w^{k-1}}{\lambda_k}, w^k - y^{k-1}\right\rangle\\ &\leq \lambda_k\alpha_k\big[\varphi_k(y^{k-1}) - \varphi_k(w^k)\big] + \lambda_k\varepsilon_k + \lambda_k\langle\nu^k, w^k - y^{k-1}\rangle.\end{aligned}$$

Denoting $\delta = \text{diam}(S_\infty)$ we get $\langle \nu^k, w^k - y^{k-1}\rangle \le \|\nu^k\|[\|w^k - y^k\| + \delta] \le \|\nu^k\|[\theta^k + \frac{1}{2} + \delta]$, so that letting $\eta_k = \lambda_k[\varepsilon_k + (\frac{1}{2} + \delta)\|\nu^k\|]$ and $\rho_k = 1 - \lambda_k\|\nu^k\|$ we obtain

$$\rho_k\theta_k \le \theta_{k-1} + \eta_k + \lambda_k\alpha_k\big[v_k - \varphi_k(w^k)\big].$$

Take $\bar{k}$ with $\rho_k > 0$ for $k \ge \bar{k}$, and let $\rho'_k = \prod_{\bar{k}}^k \rho_i$ and $\theta'_k = \rho'_k[\theta_k + \sum_{k+1}^\infty \eta_i]$. Then

$$\theta'_k \le \theta'_{k-1} + \rho'_{k-1}\lambda_k\alpha_k\big[v_k - \varphi_k(w^k)\big]. \tag{3}$$

We remark that $\sum[\varepsilon_k + \|\nu^k\|]\lambda_k < \infty$ implies $\sum_{k+1}^\infty \eta_i \to 0$ and $\rho'_k \downarrow \bar{\rho}$ for some $\bar{\rho} \in (0, 1)$.

Let $h'_k = \rho'_k[h_k + \sum_{k+1}^\infty \eta_i]$ with $h_k = \sup_z \left\{\frac{1}{2}\text{dist}(z, S_\infty)^2 : \varphi_k(z) \le v_k\right\}$. If $\theta'_k > h'_k$ then $\theta_k > h_k$ so that $\varphi_k(w^k) > v_k$ and (3) implies $\theta'_k \le \theta'_{k-1}$. Thus $\theta'_k \le \max\{h'_k, \theta'_{k-1}\}$.

We claim that $h_k \to 0$. Indeed, otherwise we could find $\epsilon > 0$, $k_j \to \infty$ and z_j with $\varphi_{k_j}(z_j) \le v_{k_j}$ and $\text{dist}(z_j, S_\infty) \ge \epsilon$. Redefining z_j if necessary (as a convex combination between z_j and its projection onto S_∞) we may suppose $\text{dist}(z_j, S_\infty) = \epsilon$. Then, by virtue of (H1) the sequence z_j is bounded, and taking subsequences we may assume it converges to some z_∞. Using (H2), (H3) and the inequality $\varphi_{k_j}(z_j) \le v_{k_j}$ we get $z_\infty \in S_\infty$ contradicting $\text{dist}(z_j, S_\infty) = \epsilon > 0$.

Since $h_k \to 0$ it follows that $h'_k \to 0$ and Lemma 1 below implies that θ'_k converges, hence θ_k converges too. We show that $\theta_{k_j} \to 0$ for some sequence $k_j \to \infty$, so that $\theta_k \to 0$ as was to be proved. Indeed, summing (3) we get $\sum[\varphi_k(w^k) - v_k]\rho'_{k-1}\alpha_k\lambda_k < \infty$ and since $\sum \rho'_{k-1}\alpha_k\lambda_k = \infty$ it follows that $\liminf_{k\to\infty}[\varphi_k(w^k) - v_k] \le 0$, so we may find $k_j \to \infty$ with $\liminf_j \varphi_{k_j}(w^{k_j}) \le v_\infty$. Since θ_k converges and S_∞ is bounded we have w^{k_j} bounded and taking a subsequence we may assume $w^{k_j} \to w^\infty$ for some w^∞. Using (H2) we get $w^\infty \in S_\infty$ so that $\theta_{k_j} \le \frac{1}{2}\|w^{k_j} - w^\infty\|^2$ and then $\theta_{k_j} \to 0$.

□

The following result was used in the previous proof, and will also be used later on.

Lemma 1. *Let θ_k, h_k, ζ_k be non-negative sequences with $\theta_k \le \max\{h_k, \theta_{k-1} + \zeta_k\}$. If $h_k \to 0$ and $\sum \zeta_k < \infty$, then θ_k converges.*

Proof. Letting $\beta_k = \theta_k + \sum_{k+1}^\infty \zeta_i$ and $\alpha_k = h_k + \sum_{k+1}^\infty \zeta_i$, we have $\beta_k \le \max\{\alpha_k, \beta_{k-1}\}$. It follows that the sequence $\gamma_k = \sup\{\beta_k, \alpha_{k+1}, \alpha_{k+2}, \ldots\}$ is non-increasing and non-negative, therefore convergent to some limit $\gamma \ge 0$. Since $\sum_{k+1}^\infty \zeta_i \to 0$ we have $\alpha_k \to 0$, and then it follows that $\beta_k \to \gamma$ from which we get $\theta_k \to \gamma$.

□

2.2. Proof of Theorem 1

In the sequel, we suppose that we are under the assumptions of Theorem 1. It is a simple matter to verify that conditions (H1), (H2) and (H3) are satisfied by the family $\varphi_k(x) = f(x, r_k)$ and $\varphi_\infty(\cdot)$ defined as $\varphi_\infty(x) = c^{\mathrm{T}}x$ if $Ax \le b$ and ∞ otherwise. Since $\sum \lambda_k = \infty$ and u^k satisfies

$$\frac{u^k - u^{k-1}}{\lambda_k} \in -\partial_{\varepsilon_k}\varphi_k(u^k) + \nu^k, \quad k = 1, 2, \ldots$$

we can apply Theorem 3 to $w^k = u^k$ with $\alpha_k \equiv 1$ and $S_\infty = S(P)$, to conclude that $\mathrm{dist}(u^k, S(P)) \to 0$. In particular, if $S(P)$ is a singleton then u^k converges to this unique optimal solution. To establish convergence of u^k when (P) admits a multiplicity of optimal solutions, we adapt the projection-scaling technique developed in [6].

Let $F = \{d \in \mathbb{R}^n : d^{\mathrm{T}} u^k \text{ converges}\}$ and let v^k and w^k be the orthogonal projections of u^k onto F and $E = F^\perp$ respectively. Clearly v^k converges, so we must only prove the convergence of w^k. Moreover, the only interesting case is when E is nontrivial. Since $S(P)$ is bounded we have $\mathrm{Ker}(A) = \{0\}$ and therefore $\mathbb{R}^n = \mathrm{span}\{A_i^{\mathrm{T}} : i = 1, \dots, m\}$, so that nontriviality of E amounts to the set $J = \{i : A_i^{\mathrm{T}} \notin F\}$ being nonempty, which we assume from now on.

Set $s^k = -(u^k - u^{k-1})/\lambda_k$. By definition of ε-subdifferential for all $x \in \mathbb{R}^n$ we have $\varphi_k(u^k) + \langle s^k + \nu^k, x - u^k\rangle \le \varphi_k(x) + \varepsilon_k$. Taking $x = w + v^k$ with $w \in E$ we obtain

$$\varphi_k(w^k + v^k) + \langle s^k + \nu^k, w - w^k\rangle \le \varphi_k(w + v^k) + \varepsilon_k$$

and denoting $\tilde{s}^k$ and $\tilde{\nu}^k$ the projections of s^k and ν^k onto the space E we deduce $\tilde{s}^k + \tilde{\nu}^k \in \partial_{\varepsilon_k} \tilde{\varphi}_k(w^k)$ with $\tilde{\varphi}_k : E \to \mathbb{R}$ given by $\tilde{\varphi}_k(w) = \varphi_k(w + v^k)$. Observing that $c \in F$ (since $c^{\mathrm{T}} u^k$ converges to the optimal value of (P)) and $A_i \in F$ for $i \notin J$, we get

$$\tilde{\varphi}_k(w) = c^{\mathrm{T}} v^k + r_k \textstyle\sum_{i \notin J} \exp\left(-b_i^k/r_k\right) + r_k \sum_{i \in J} \exp\left(\left[A_i w - b_i^k\right]/r_k\right)$$

with $b_i^k = b_i - A_i v^k$ which converges to some b_i^∞ as $k \to \infty$. Then, considering the auxiliary family of convex functions $\hat{\varphi}_k \in \Gamma_0(E)$ defined by

$$\hat{\varphi}_k(w) = r_k \ln\left(\textstyle\sum_{i \in J} \exp\left(\left[A_i w - b_i^k\right]/r_k\right)\right).$$

and setting $g_k(t) = r_k \exp(t/r_k)$ we have $\tilde{\varphi}_k(w) = g_k \circ \hat{\varphi}_k(w) + C_k$ for a suitable constant $C_k \in \mathbb{R}$, so that $\tilde{s}^k + \tilde{\nu}^k \in \partial_{\varepsilon_k}(g_k \circ \hat{\varphi}_k)(w^k)$. By [11, Theorem 3.6.1], there exist $\varepsilon_k^1 \ge 0$, $\varepsilon_k^2 \ge 0$ with $\varepsilon_k^1 + \varepsilon_k^2 = \varepsilon_k$ and a scalar factor $\alpha_k \in \partial_{\varepsilon_k^2} g_k(\hat{\varphi}_k(w^k))$ such that $\tilde{s}^k + \tilde{\nu}^k \in \partial_{\varepsilon_k^1}(\alpha_k \hat{\varphi}_k)(w^k) \subset \partial_{\varepsilon_k}(\alpha_k \hat{\varphi}_k)(w^k)$. Since $\tilde{s}^k = -(w^k - w^{k-1})/\lambda_k$ we conclude

$$\frac{w^k - w^{k-1}}{\lambda_k} \in -\partial_{\varepsilon_k}(\alpha_k \hat{\varphi}_k)(w^k) + \tilde{\nu}^k \tag{4}$$

with $\alpha_k \ge 0$ (because g_k is increasing) and $\sum[\varepsilon_k + \|\tilde{\nu}^k\|]\lambda_k < \infty$.

Now $\hat{\varphi}_k(\cdot)$ is L-Lipschitz with $L = \max_{i \in J} \|A_i\|$ so that (4) implies $\|w^k - w^{k-1}\| \le [L\alpha_k + \|\tilde{\nu}^k\|]\lambda_k$. We claim that $\sum \alpha_k \lambda_k < \infty$ from which it follows that w^k is Cauchy, hence convergent, so that u^k converges which completes the proof.

To prove this claim we proceed by contradiction. Suppose $\sum \alpha_k \lambda_k = \infty$ and let us apply Theorem 3 to (4) with $\hat{\varphi}_\infty(w) = \max_{i \in J}\{A_i w - b_i^\infty\}$. The boundedness of $S(P)$ implies that $\hat{\varphi}_\infty$ is coercive, and therefore $\hat{v}_\infty = \inf_{w \in E} \hat{\varphi}_\infty(w)$ is finite with $\hat{S}_\infty = \mathrm{Argmin}_{w \in E} \hat{\varphi}_\infty(w)$ nonempty and bounded, proving (H1). Verification of (H2) and (H3) is immediate from

$$\max_{i \in J}\left\{A_i w - b_i^k\right\} \le \hat{\varphi}_k(w) \le \max_{i \in J}\left\{A_i w - b_i^k\right\} + r_k \ln|J| \quad \forall w \in E,$$

so we may use Theorem 3 to deduce that $\lim_{k\to\infty} \mathrm{dist}(w^k, \hat{S}_\infty) = 0$. Choose $i_0 \in J$ such that $A_{i_0} w - b_{i_0}^\infty = \hat{v}_\infty$ for all $w \in \hat{S}_\infty$ (existence of i_0 follows by contradiction: otherwise for each $i \in J$ we could find $w_i \in \hat{S}_\infty$ with $A_i w_i - b_i^\infty < \hat{v}_\infty$, and the point $\bar{w} = \frac{1}{|J|}\sum_{i\in J} w_i \in \hat{S}_\infty$ would satisfy $\hat{\varphi}_\infty(\bar{w}) < \hat{v}_\infty$ which is absurd). Then $\mathrm{dist}(w^k, \hat{S}_\infty) \to 0$ implies that $A_{i_0} w^k \to b_{i_0}^\infty + \hat{v}_\infty$, and therefore $A_{i_0} u^k$ converges. This amounts to say that $A_{i_0} \in F$ contradicting the fact that $i_0 \in J$. This contradiction proves our claim $\sum \alpha_k \lambda_k < \infty$ and completes the proof.

□

3. Proof of Theorem 2: dual convergence

Let $\psi_k = \sum_{i=1}^m \mu_i^k - \lambda_i^* \ln \mu_i^k$ and $\alpha = \sum_{i=1}^m \lambda_i^* - \lambda_i^* \ln \lambda_i^*$. Since $\mu_i \mapsto \mu_i - \lambda_i^* \ln \mu_i$ attains its minimum on $[0, \infty)$ at the unique point $\mu_i = \lambda_i^*$ it follows that $\psi_k \geq \alpha$, and $\mu^k \to \lambda^*$ if and only if $\psi_k \to \alpha$. Hence the result boils down to showing that $\theta_k \to 0$ with $\theta_k = \psi_k - \alpha \geq 0$.

Let us first introduce some additional notation. We denote $s^k = -(u^k - u^{k-1})/\lambda_k$, $g^k = s^k + v^k \in \partial_{\varepsilon_k} f(u^k, r_k)$ and $\eta^k = v^k + q_k(u^{k-1} - u^\infty)$ where $u^\infty = \lim_{k\to\infty} u^k$. We also set $I_\infty = \{i : A_i u^\infty = b_i\}$ and $\beta_k = \sqrt{\varepsilon_k/r_k} + \varepsilon_k/r_k + \sum_{i\notin I_\infty} \mu_i^k$. Notice that assumptions (d), (e) and (f) imply that $\eta^k \to 0$ and $\beta_k \to 0$. To proceed with the proof we require the following technical lemmas (to be proved later on).

Lemma 2. $\theta_k - \theta_{k-1} \leq [\varepsilon_k - \lambda_k \langle g^k, g^k - \eta^k\rangle]/r_k$.

Lemma 3. *There exists a nondecreasing function* $\rho : [0, \infty) \to \mathbb{R}$ *such that* $\lim_{t\to 0} \rho(t) = 0$ *and* $\theta_k \leq \rho(\|g^k\| + \beta_k)$.

The proof is established in the following two claims.

Claim 1. θ_k converges.

We distinguish the alternative cases in (g).

If $\varepsilon_k/\lambda_k \to 0$ from Lemma 2 we see that $\theta_k > \theta_{k-1}$ implies $\|g^k\| \leq \gamma_k$ with $\gamma_k = \frac{1}{2}[\|\eta^k\| + \sqrt{\|\eta^k\|^2 + 4\varepsilon_k/\lambda_k}] \to 0$, so that Lemma 3 yields $\theta_k \leq h_k := \rho(\gamma_k + \beta_k)$ and then $\theta_k \leq \max\{h_k, \theta_{k-1}\}$. Since $h_k \to 0$ the claim follows from Lemma 1.

The alternative case $\sum \varepsilon_k/r_k < \infty$ is similar: if $\|g^k\| > \|\eta^k\|$ then Lemma 2 gives $\theta_k \leq \theta_{k-1} + \varepsilon_k/r_k$, while $\|g^k\| \leq \|\eta^k\|$ implies by Lemma 3 that $\theta_k \leq h_k := \rho(\|\eta^k\| + \beta_k)$. Hence $\theta_k \leq \max\{h_k, \theta_{k-1} + \varepsilon_k/r_k\}$ and the claim follows again from Lemma 1.

Claim 2. $\theta_k \to 0$.

Using Claim 1 and Lemma 3 it suffices to prove that $\liminf \|g^k\| = 0$. If this was not the case we could find $c > 0$ with $\langle g^k, g^k - \eta^k\rangle > c$ for all k large, and Lemma 2 would imply $[c\lambda_k - \varepsilon_k]/r_k \leq \theta_{k-1} - \theta_k$. Considering the alternative cases in (g) we get $\sum \lambda_k/r_k < \infty$ which yields a contradiction.

□

Let us now prove the technical Lemmas 2 and 3.

Proof of Lemma 2. Let us observe that $\psi_k = [f(u^k, r_k) - c^{\mathrm{T}} u^\infty]/r_k$. Letting $x^k = (1-\delta_k)u^\infty + \delta_k u^{k-1}$ with $\delta_k = r_k/r_{k-1} \in [0,1]$, and since $A_i u^\infty \le b_i$, we have $A_i x^k - b_i \le \delta_k(A_i u^{k-1} - b_i)$ and then

$$\begin{aligned} f(x^k, r_k) &\le (1-\delta_k)c^{\mathrm{T}}u^\infty + \delta_k c^{\mathrm{T}} u^{k-1} + r_k \textstyle\sum_{i=1}^m \exp\left(\delta_k\left[A_i u^{k-1} - b_i\right]/r_k\right) \qquad (5)\\ &= c^{\mathrm{T}}u^\infty + \delta_k\left[f\left(u^{k-1}, r_{k-1}\right) - c^{\mathrm{T}}u^\infty\right] \\ &= c^{\mathrm{T}}u^\infty + r_k\psi_{k-1}. \end{aligned}$$

Now, taking $x = x^k$ in the subgradient inequality

$$f\left(u^k, r_k\right) + \langle g^k, x - u^k\rangle \le f(x, r_k) + \varepsilon_k \qquad (6)$$

and combining with (5) we obtain

$$f\left(u^k, r_k\right) - c^{\mathrm{T}}u^\infty + \langle g^k, x^k - u^k\rangle \le r_k\psi_{k-1} + \varepsilon_k.$$

Since $x^k - u^k = \lambda_k(g^k - \eta^k)$, dividing by r_k we get $\psi_k + \lambda_k\langle g^k, g^k - \eta^k\rangle/r_k \le \psi_{k-1} + \varepsilon_k/r_k$ from which the result follows at once.

□

In order to prove Lemma 3 we need a bound on the multipliers μ_i^k for $i \notin I_0$, where we denote $I_0 = \{i : A_i x = b_i \text{ for all } x \in S(P)\}$.

Lemma 4. *There exists $C \ge 0$ such that $\sum_{i\notin I_0} \mu_i^k \le C[\|g^k\| + \beta_k]$.*

Proof. Let x^* be the centroid or the optimal face of (P). Taking $x = u^k + r_k(x^* - u^\infty)$ in the subgradient inequality (6) we obtain

$$\sum_{i=1}^m \mu_i^k + \langle g^k, x^* - u^\infty\rangle \le \sum_{i=1}^m \mu_i^k \exp\left(A_i\left[x^* - u^\infty\right]\right) + \varepsilon_k/r_k.$$

Since for $i \in I_0$, $A_i x^* = A_i u^\infty = b_i$, we deduce

$$\begin{aligned} \sum_{i\notin I_0} \mu_i^k &\le \|g^k\|\|u^\infty - x^*\| + \sum_{i\notin I_0} \mu_i^k \exp\left(A_i\left[x^* - u^\infty\right]\right) + \varepsilon_k/r_k \\ &\le \|g^k\|\|u^\infty - x^*\| + \sum_{i\in I_\infty\setminus I_0} \mu_i^k \exp(A_i x^* - b_i) + L\sum_{i\notin I_\infty} \mu_i^k + \varepsilon_k/r_k \end{aligned}$$

with $L = \max_{i\notin I_\infty} \exp(A_i x^* - A_i u^\infty)$, and hence

$$\sum_{i\in I_\infty\setminus I_0} \mu_i^k[1 - \exp(A_i x^* - b_i)] \le \|g^k\|\|u^\infty - x^*\| + L\sum_{i\notin I_\infty} \mu_i^k + \varepsilon_k/r_k.$$

Since $A_i x^* < b_i$ for all $i \notin I_0$ (see [8]), we have $\exp(A_i x^* - b_i) < 1$ for $i \notin I_0$, and therefore we may find a constant $L' \ge 0$ such that

$$\sum_{i\in I_\infty\setminus I_0} \mu_i^k \le L'\left[\|g^k\| + \sum_{i\notin I_\infty} \mu_i^k + \varepsilon_k/r_k\right] \le L'\left[\|g^k\| + \beta_k\right]$$

so that the Lemma holds with $C = L' + 1$.

□

Proof of Lemma 3. From (6) it follows that for every $x \in \mathbb{R}^n$

$$\left[f\left(u^k, r_k\right) - c^{\mathrm{T}}u^\infty\right]/r_k + \left\langle g^k, (x - u^k)/r_k\right\rangle \leq \left[f(x, r_k) - c^{\mathrm{T}}u^\infty\right]/r_k + \varepsilon_k/r_k.$$

Taking $x = u^\infty + r_k z$ and letting $\varphi_k(z) = c^{\mathrm{T}}z + \sum_{i=1}^m \exp(A_i z)\exp[(A_i u^\infty - b_i)/r_k]$ and $z^k = (u^k - u^\infty)/r_k$, we get

$$\varphi_k(z^k) + \langle g^k, z - z^k\rangle \leq \varphi_k(z) + \varepsilon_k/r_k$$

so that $g^k \in \partial_{\varepsilon_k/r_k}\varphi_k(z^k)$. Using Brondsted-Rockafellar's theorem [11, Theorem 4.2.1], we may find $\tilde{z}^k \in \mathbb{R}^n$ such that $\|z^k - \tilde{z}^k\| \leq \sqrt{\varepsilon_k/r_k}$ and $\|g^k - \nabla\varphi_k(\tilde{z}^k)\| \leq \sqrt{\varepsilon_k/r_k}$. A direct computation gives

$$\begin{aligned}\left\|c + \textstyle\sum_{i\in I_0} \exp\left(A_i\tilde{z}^k\right)A_i^{\mathrm{T}}\right\| &\leq \sqrt{\varepsilon_k/r_k} + \|g^k\| + \textstyle\sum_{i\notin I_0} \mu_i^k \exp\left(A_i[\tilde{z}^k - z^k]\right)\left\|A_i^{\mathrm{T}}\right\| \\ &\leq \sqrt{\varepsilon_k/r_k} + \|g^k\| + C_1\exp(C_1\sqrt{\varepsilon_k/r_k})\textstyle\sum_{i\notin I_0}\mu_i^k\end{aligned}$$

with $C_1 = \max_{i\notin I_0}\|A_i^{\mathrm{T}}\|$, and then Lemma 4 implies that for some $C' > 0$ we have

$$\left\|c + \textstyle\sum_{i\in I_0} \exp\left(A_i\tilde{z}^k\right)A_i^{\mathrm{T}}\right\| \leq C'\left[\|g^k\| + \beta_k\right]. \tag{7}$$

Let $E_0 = \mathrm{span}\{A_i : i \in I_0\}$ and define $\varphi : E_0 \to \mathbb{R}$ by $\varphi(v) = c^{\mathrm{T}}v + \sum_{i\in I_0}\exp(A_i v)$. Letting $\tilde{v}^k$ be the projection of $\tilde{z}^k$ onto E_0, (7) becomes $\|\nabla\varphi(\tilde{v}^k)\| \leq C'[\|g^k\| + \beta_k]$. It is known (cf. [8, p. 180]) that φ is strictly convex and coercive with its minimum equal to α and attained at a unique point $v_0 \in E_0$. Consequently, the functions

$$\begin{aligned} g(t) &= \sup\{\varphi(v) : \|v - v_0\| \leq t\} \\ h(t) &= \sup\{\|v - v_0\| : \|\nabla\varphi(v)\| \leq t\}\end{aligned}$$

are non-decreasing and converge as $t \to 0$ towards α and 0 respectively. Letting v^k be the projection of z^k onto E_0, we have $\|v^k - \tilde{v}^k\| \leq \sqrt{\varepsilon_k/r_k}$ and therefore

$$\varphi(v^k) \leq g\left(\sqrt{\varepsilon_k/r_k} + \left\|\tilde{v}^k - v_0\right\|\right) \leq g\left(\beta_k + h\left(C'\left[\|g^k\| + \beta_k\right]\right)\right).$$

Since $\psi_k = \varphi(v^k) + \sum_{i\notin I_0}\mu_i^k$, using Lemma 4 we get

$$\psi_k \leq g\left(\beta_k + h\left(C'\left[\|g^k\| + \beta_k\right]\right)\right) + C\left(\|g^k\| + \beta_k\right)$$

and then Lemma 3 holds with $\rho(t) = g(t + h(C't)) + Ct - \alpha$.

□

Acknowledgement. We thank an anonymous referee for suggesting the possibility of including a residual ν^k in the solution of (PE). We are also indebted to Matías Courdurier for useful discussions on this question.

References

1. Alart, P., Lemaire, B. (1991): Penalization in nonclassical convex programming via variational convergence. Math. Program. **51**, 307–331
2. Alvarez, F. (2000): Absolute minimizer in convex programming by exponential penalty. J. Convex Anal. **7**, 197–202
3. Auslender, A. (1987): Numerical methods for nondifferentiable convex optimization. Math. Program. Studies **30**, 102–127
4. Auslender, A., Crouzeix, J.P., Fedit, P. (1987): Penalty-proximal methods in convex programming. J. Optim. Theory Appl. **55**, 1–21
5. Bahraoui, M.A., Lemaire, B. (1994): Convergence of diagonally stationary sequences in convex optimization. Set-Valued Analysis **2**, 49–61
6. Baillon, J.B., Cominetti, R. (2001): A convergence result for non-autonomous subgradient evolution equations and its application to the steepest descent exponential penalty trajectory in linear programming. J. Funct. Anal. **187**, 263–273
7. Cominetti, R. (1997): Coupling the proximal point algorithm with approximation methods. J. Optim. Theory Appl. **95**, 581–600
8. Cominetti, R., San Martín, J. (1994): Trajectory analysis for the exponential penalty method in linear programming. Math. Program. **67**, 169–187
9. Correa, R., Lemaréchal, C. (1993): Convergence of some algorithms for convex optimization. Math. Program. **62**, 261–275
10. Güler, O. (1991): On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403–419
11. Hiriart-Urruty, J.B., Lemaréchal, C. (1996): Convex Analysis and Minimization Algorithms II. Springer, Berlin
12. Kaplan, A. (1978): On a convex programming method with internal regularization. Soviet Math. Doklady **19**, 795–799
13. Kaplan, A., Tichatschke, R. (1998): Proximal methods in view of interior-point strategies. J. Optim. Theory Appl. **98**(2), 399–429
14. Lemaire, B. (1988): Coupling optimization methods and variational convergence. Trends in Mathematical Optimization, International Series of Numerical Mathematics, Birkhäuser, Basel, pp. 163–179
15. Lemaire, B. (1992): About the convergence of the proximal method. Advances in Optimization Lecture Notes in Economics and Mathematical Systems 382, pp. 39–51. Springer, Berlin
16. Martinet, B. (1970): Régularisation d'inéquations variationnelles par approximations successives. Revue Française d'Informatique et Recherche Operationnelle **4**, 154–159
17. Mouallif, K. (1987): Sur la convergence d'une méthode associant pénalisation et réguralisation. Bulletin de la Societé Royale des Sciences de Liège **56**, 175–180
18. Mouallif, K., Tossings, P. (1987): Une méthode de pénalisation exponentielle associée à une régularisation proximale. Bulletin de la Societé Royale des Sciences de Liège **56**, 181–190
19. Mouallif, K., Tossings, P. (1990): Variational metric and exponential penalization. J. Optim. Theory Appl. **67**(1), 185–192
20. Moudafi, A. (1994): Coupling proximal algorithm and Tikhonov method. Nonlinear Times and Digest **1**, 203–210
21. Rockafellar, R.T. (1976): Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898