SOLVING LARGE-SCALE MINIMAX PROBLEMS WITH THE PRIMAL-DUAL STEEPEST DESCENT ALGORITHM

Ciyou Zhu *

Department of Mathematical Sciences
Johns Hopkins University, Baltimore, MD 21218

June 1992; revised June 1993.

Abstract. This paper shows that the primal-dual steepest descent algorithm developed Zhu and Rockafellar for large-scale extended linear-quadratic programming can be used in solving constrained minimax problems related to a general C^2 saddle function. It is proved that the algorithm converges linearly from the very beginning of the iteration if the related saddle function is strongly convex-concave uniformly and the cross elements between the convex part and the concave part of the variables in its Hessian are bounded on the feasible region. Better bounds for the asymptotic rates of convergence are also obtained. The minimax problems where the saddle function has linear cross terms between the convex part and the concave part of the variables are discussed specifically as a generalization of the extended linear-quadratic programming. Some fundamental features of these problems are laid out and analyzed.

Keywords. Minimax problem, saddle function, large-scale numerical optimization, primal-dual projected gradient algorithm.

---

* This work was supported by Eliezer Naddor Postdoctoral Fellowship in Mathematical Sciences at the Department of Mathematical Sciences, the Johns Hopkins University, during the year 1991–92.

Currrent address: Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439.

1. Introduction.

Let U and V be nonempty closed convex sets in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, and let the Lagrangian $L(u,v)$ defined on $U\times V$ be a finite-valued saddle function convex in u and concave in v. Consider the problem of finding a saddle point $(\bar u,\bar v)$ of $L(u,v)$ over $U\times V$, i.e., finding a pair $(\bar u,\bar v)\in U\times V$ such that

$$L(\bar u,v)\le L(\bar u,\bar v)\le L(u,\bar v)\quad \forall u\in U,\,\forall v\in V. \tag{1.1}$$

As is well known, a large variety of optimization problems can be cast in this formulation. For instance, the convex programming problem

$$\begin{aligned}&\text{minimize } \varphi(u) \text{ over } u\in U,\\ &\text{subject to } h_i(u)\le 0,\ i=1,\dots,m,\end{aligned}$$

with φ and h_i's being finite convex functions on U, is equivalent to the minimax problem of the Lagrangian

$$L(u,v)=\varphi(u)+\sum_{i=1}^{m} v_i h_i(u) \text{ over } U\times\mathbb{R}^m_+.$$

Another special case which has significant large-scale application is the extended linear-quadratic programming problems (ELQP for short) recently introduced in the context of multistage and stochastic optimization [3–11], where the sets U and V are polyhedral, and the associated saddle function $L(u,v)$ is linear-quadratic,

$$L(u,v)=p{\cdot}u+\tfrac12 u{\cdot}Pu+q{\cdot}v-\tfrac12 v{\cdot}Qv-v{\cdot}Ru, \tag{1.2}$$

with the matrices $P\in\mathbb{R}^{n\times n}$ and $Q\in\mathbb{R}^{m\times m}$ being symmetric and positive semidefinite. (One has $p\in\mathbb{R}^n$, $q\in\mathbb{R}^m$, and $R\in\mathbb{R}^{m\times n}$.) In Section 5, we generalize ELQP to the minimax problem where the Lagrangian has linear cross terms between the variables u and v. The latter serves as an example to which our algorithm applies, as well as one with significant large-scale applications.

Associated with L, U and V are the primal and dual problems

$(\mathcal{P})$ minimize $f(u)$ over all $u\in U$, where $f(u):=\sup_{v\in V} L(u,v)$,

$(\mathcal{Q})$ maximize $g(v)$ over all $v\in V$, where $g(v):=\inf_{u\in U} L(u,v)$.

The relationship between $(\mathcal{P})$ and $(\mathcal{Q})$ is included in the following theorem.

Theorem 1.1 [12, Theorem 36.2] (duality and optimality). *A point* $(\bar{u}, \bar{v})$ *is a saddle point of* L *over* $U \times V$ *if and only if* $\inf(\mathcal{P})$ *is attained at* $\bar{u}$, $\sup(\mathcal{Q})$ *is attained at* $\bar{v}$, *and these two extrema are equal. If* $(\bar{u}, \bar{v})$ *is a saddle point of* $L(u, v)$ *over* $U \times V$, *then* $f(\bar{u}) = L(\bar{u}, \bar{v}) = g(\bar{v})$.

Let $\delta_S(\cdot)$ be the indicator function related to a convex set S, and let $\operatorname{ri} S$ be the relative interior of S. The results on the existence of a saddle point may be written in the following form.

Theorem 1.2 [12, Theorem 37.6] (existence of saddle points). *If both of the following conditions hold, then there exists a saddle point* $(\bar{u}, \bar{v})$ *of* $L(u, v)$ *over* $U \times V$.

(a) *The convex functions* $L(\cdot, v) + \delta_U(\cdot)$ *for* $v \in \operatorname{ri} V$ *have no common direction of recession.*

(b) *The convex functions* $-L(u, \cdot) + \delta_V(\cdot)$ *for* $u \in \operatorname{ri} U$ *have no common direction of recession.*

It is easy to see that condition (a) in Theorem 1.2 will be satisfied if either $L(u, v)$ is strongly convex in u or the set U is bounded. Similarly condition (b) will be satisfied if either $L(u, v)$ is strongly concave in v or the set V is bounded. In the case of minimax problems where the Lagrangian has linear cross terms, less stringent criteria for these two conditions to be satisfied will be given in Section 5.

For solving the optimization problems posed in this minimax setting, one possible approach, which is capable of taking advantage of the problem structure in large-scale applications, is the splitting method. See Chen and Rockafellar [13] for a thorough discussion on the forward-backward splitting methods in Lagrangian optimization, and Zhu [14] for some recent convergence results. In this paper, we are going to show that another possible approach, which can also take advantage of the problem structure and has been tested successfully on large-scale ELQP, is the primal-dual steepest descent algorithm developed in [1, 2].

The primal-dual steepest descent algorithm (PDSD for short) was proposed by Zhu and Rockafellar [1] as a method designed for solving large-scale ELQP problems arising in dynamic and stochastic optimization. The name stems from the interpretation of the search direction as corresponding to the projected gradient on the feasible set in the norm induced by a certain matrix in the Lagrangian. This algorithm shows promise for large-scale optimization problems since it requires relatively simple computations to be performed in a manner conducive to decomposition.

In [2], Zhu developed new variants of PDSD for ELQP by introducing different update schemes and step length rules. All the variants were put in a unified framework. Zhu proved that the algorithm, when applied to ELQP, converges linearly from the very beginning of the iteration. New estimations for the convergence ratio of the algorithm were also obtained.

In this paper, we extend the algorithm and its convergence theory to the class of minimax problems related to a general C^2 saddle function. In Section 2, we present the algorithm after laying out some computationally useful properties of the problem. These properties are the counterparts of the corresponding results of Rockafellar [8] on ELQP. In Section 3, we prove global linear convergence results for each variant of the algorithm formulated in Section 2. In Section 4, we derive estimates for the asymptotic rates of convergence of the algorithm. Finally, in Section 5, we discuss the minimax problem with linear cross terms between the variables u and v in the Lagrangian as a typical example fitting the assumptions of the paper. We shall use the Euclidean inner product and norm throughout the paper, denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ respectively. We use $[w_1, w_2]$ to denote the line segment between two points w_1 and w_2, and use $]w_1, w_2[$ to denote the same line segment with the endpoints excluded. The meanings of $[w_1, w_2[$ and $]w_1, w_2]$ are self-evident from the above conventions.

2. The Primal-Dual Steepest Descent Algorithm.

In this section, we formulate six variants of the algorithm in a unified framework. We first derive some computationally useful properties of the general minimax problem. Suppose the Lagrangian $L(u, v)$ is C^2 on some open set containing $U \times V$. For any $U' \subset U$ and $V' \subset V$ define

$$\lambda_{U',V'} := \inf\{ \text{ smallest eigenvalue of } \nabla^2_{uu} L(u, v) \mid u \in U', v \in V'\} \qquad (2.1a)$$

$$\lambda_{V',U'} := \inf\{ \text{ smallest eigenvalue of } -\nabla^2_{vv} L(u, v) \mid u \in U', v \in V'\} \qquad (2.1b)$$

$$M_{U',V'} := \sup\{\|\nabla^2_{uv} L(u, v)\| \mid u \in U', v \in V'\}. \qquad (2.1c)$$

Note that the order of the subscripts of λ also indicates with respect to which variable the differentiation is performed. The convexity-concavity of L implies $\lambda_{U,V} \geq 0$

and $\lambda_{V,U} \geq 0$. From now on, we focus on the general minimax problems with a strongly convex-concave Lagrangian. (If the Lagrangian of the original problem does not exhibit this property, the proximal point algorithm can be implemented as an "outer" loop of iteration [8, 15] to create a strongly convex-concave Lagrangian for the "inner" loop problems.) The following assumption is imposed on the rest of the paper.

Assumption 2.1 (blanket assumption)**.** *$M_{U,V} < +\infty$ and the Lagrangian $L(u,v)$ is strongly convex-concave uniformly on $U \times V$ in the sense that $\lambda_{U,V} > 0$ and $\lambda_{V,U} > 0$.*

Under this assumption, the subproblems of maximizing $L(u,v)$ in v for fixed u or minimizing $L(u,v)$ in u for fixed v calculate not only the objective values $f(u)$ and $g(v)$, but also the unique vectors

$$F(u) = \operatorname*{argmax}_{v \in V} L(u,v) \quad \text{and} \quad G(v) = \operatorname*{argmin}_{u \in U} L(u,v). \tag{2.2}$$

For large-scale problems, we say that the minimax problem possesses *double decomposability* (or Lagrangian decomposability) [8] if these subproblems can be decomposed to low dimensional ones and solved easily. The *box-separable* case of the minimax problem with LCT, which will be discussed in Section 5, is a typical example of that kind. We shall refer consistently to

$$\begin{aligned} \bar{u} = &\ \text{the unique optimal solution to } (\mathcal{P}), \\ \bar{v} = &\ \text{the unique optimal solution to } (\mathcal{Q}). \end{aligned}$$

The next two propositions are generalizations of the corresponding results on ELQP in Rockafellar [8].

Proposition 2.2 (optimality estimates)**.** *Suppose $\hat{u}$ and $\hat{v}$ are elements of U and V satisfying $f(\hat{u}) - g(\hat{v}) \leq \varepsilon$ for a certain $\varepsilon \geq 0$. Then $\hat{u}$ and $\hat{v}$ are ε-optimal in the sense that $|f(\hat{u}) - f(\bar{u})| \leq \varepsilon$ and $|g(\hat{v}) - g(\bar{v})| \leq \varepsilon$. Moreover,*

$$\lambda_{U,V}\|\hat{u} - \bar{u}\|^2 + \lambda_{V,U}\|\hat{v} - \bar{v}\|^2 \leq 2\varepsilon. \tag{2.3}$$

Proof. By Theorem 1.1,

$$f(\hat{u}) \geq f(\bar{u}) = L(\bar{u},\bar{v}) = g(\bar{v}) \geq g(\hat{v}),$$

from which the inequalities for f and g follows. Now $f(\hat{u}) \geq L(\hat{u}, \bar{v})$ by the definition of f. Hence

$$\begin{aligned} f(\hat{u}) - L(\bar{u}, \bar{v}) &\geq L(\hat{u}, \bar{v}) - L(\bar{u}, \bar{v}) \\ &= \nabla_u L(\bar{u}, \bar{v}) \cdot (\hat{u} - \bar{u}) + \tfrac{1}{2}(\hat{u} - \bar{u}) \cdot \nabla^2_{uu} L(\tilde{u}, \bar{v})(\hat{u} - \bar{u}), \end{aligned}$$

where $\tilde{u} \in [\hat{u}, \bar{u}]$. Observe that $\nabla_u L(\bar{u}, \bar{v}) \cdot (\hat{u} - \bar{u}) \geq 0$, since $\bar{u}$ minimizes $L(u, \bar{v})$ on $u \in U$. Moreover, $(\hat{u} - \bar{u}) \cdot \nabla^2_{uu} L(\tilde{u}, \bar{v})(\hat{u} - \bar{u}) \geq \lambda_{U,V} \|\hat{u} - \bar{u}\|^2$ by (2.1). Therefore

$$f(\hat{u}) - L(\bar{u}, \bar{v}) \geq \tfrac{1}{2} \lambda_{U,V} \|\hat{u} - \bar{u}\|^2. \tag{2.4}$$

Similarly, we have

$$g(\hat{v}) - L(\bar{u}, \bar{v}) \leq -\tfrac{1}{2} \lambda_{V,U} \|\hat{v} - \bar{v}\|^2. \tag{2.5}$$

Subtracting (2.5) from (2.4), we obtain (2.3). □

Proposition 2.3 (Lipschitz properties of F and G). *The mappings F and G defined by (2.2) are Lipschitz continuous on U and V respectively with*

$$\|F(u') - F(u)\| \leq \big(M_{U,V} / \lambda_{V,U}\big) \|u' - u\| \quad \forall u \in U,\ u' \in U, \tag{2.6a}$$

$$\|G(v') - G(v)\| \leq \big(M_{U,V} / \lambda_{U,V}\big) \|v' - v\| \quad \forall v \in V,\ v' \in V. \tag{2.6b}$$

Proof. For any $u_0, u_0' \in U$, let $v_1 = F(u_0)$ and $v_1' = F(u_0')$. Then it follows from the definition of F in (2.2) and the first-order optimality condition that

$$\nabla_v L(u_0, v_1) \cdot (v - v_1) \leq 0 \quad \forall v \in V \quad \text{and} \quad \nabla_v L(u_0', v_1') \cdot (v - v_1') \leq 0 \quad \forall v \in V.$$

Adding these two inequalities with $v = v_1'$ in the first and $v = v_1$ in the second, we have

$$\begin{aligned} 0 \geq& \nabla_v L(u_0, v_1) \cdot (v_1' - v_1) + \nabla_v L(u_0', v_1') \cdot (v_1 - v_1') \\ =& [\nabla_v L(u_0, v_1) - \nabla_v L(u_0', v_1')] \cdot (v_1' - v_1) \\ =& -(v_1' - v_1) \cdot \nabla^2_{vu} L(\tilde{u}, \tilde{v})(u_0' - u_0) - (v_1' - v_1) \cdot \nabla^2_{vv} L(\tilde{u}, \tilde{v})(v_1' - v_1), \end{aligned}$$

for some $\tilde{u} \in [u_0', u_0]$ and $\tilde{v} \in [v_1', v_1]$. Then

$$(v_1' - v_1) \cdot \nabla^2_{vu} L(\tilde{u}, \tilde{v})(u_0' - u_0) \geq -(v_1' - v_1) \cdot \nabla^2_{vv} L(\tilde{u}, \tilde{v})(v_1' - v_1) \geq 0.$$

Therefore with the notations defined in (2.1),

$$M_{U,V}\|v_1' - v_1\|\|u_0' - u_0\| \geq \lambda_{V,U}\|v_1' - v_1\|^2,$$

which yields (2.6a). We can prove (2.6b) similarly. $\square$

The points generated by the mappings F and G as by-products when calculating the objective values contain important information of the problem. Our effort in designing the algorithm is concentrated on using this information effectively in a large scale setting. The PDSD algorithm first searches on line segments $[u, G(F(u))]$ and $[v, F(G(v))]$ in primal and dual variables respectively to get some *intermediate points* as candidates for the next iterates. Then the mappings F and G are used again to pass information between the primal part and the dual part of the algorithm in determining the next pair of primal and dual iterates.

Define

$$\tilde{\sigma} = \frac{M_{U,V}^2}{\lambda_{U,V}\lambda_{V,U}}. \tag{2.7}$$

The following PDSD algorithm for the general minmax problem contains six variants, including three different step length rules and two update schemes. We shall refer to the algorithm with, say, update scheme 2 and step length rule (iii), as PDSD–2(iii).

Primal-Dual Steepest Descent Algorithm.

Step 0 (initialization). *Set $\nu := 0$ (iteration counter). Specify starting points $u^0 \in U$ and $v^0 \in V$. Choose one of the step length rules in Step 2. (If rule (iii) is chosen, then also choose some constant $\delta \in (0,1)$.) Choose one of the update schemes in Step 3. Construct primal and dual sequences $\{u^\nu\} \subset U$ and $\{v^\nu\} \subset V$ as follows.*

Step 1 (optimality test). *If*

$$\min\{\, f(u^\nu), f\big(G(v^\nu)\big)\,\} - \max\{\, g(v^\nu), g\big(F(u^\nu)\big)\,\} = 0,$$

then terminate with

$$\bar{u} = \operatorname{argmin}\{\, f(u) \,|\, u = u^\nu, \;\; \text{or } u = G(v^\nu)\,\}$$
$$\bar{v} = \operatorname{argmax}\{\, g(v) \,|\, v = v^\nu, \;\; \text{or } v = F(u^\nu)\,\}$$

being optimal solutions to $(\mathcal{P})$ *and* $(\mathcal{Q})$.

Step 2 (line search). *Use one of the following step length rules chosen at initialization to determine* α_ν *and* β_ν *for generating intermediate points*

$$\hat{u}^{\nu+1} :=(1-\alpha_\nu)u^\nu + \alpha_\nu G(F(u^\nu)),$$
$$\hat{v}^{\nu+1} :=(1-\beta_\nu)v^\nu + \beta_\nu F(G(v^\nu)),$$

in primal and dual variables respectively.

(i) *Perfect line search:*

$$\alpha_\nu := \operatorname*{argmin}_{\alpha\in[0,1]} f\big((1-\alpha)u^\nu + \alpha G(F(u^\nu))\big),$$
$$\beta_\nu := \operatorname*{argmax}_{\beta\in[0,1]} g\big((1-\beta)v^\nu + \beta F(G(v^\nu))\big).$$

(ii) *Fixed step lengths:*

$$\alpha_\nu := \min\{1, \frac{1}{2\tilde{\sigma}}\} \quad \text{and} \quad \beta_\nu := \min\{1, \frac{1}{2\tilde{\sigma}}\}.$$

(We adopt the convention $0^{-1} = +\infty$ *in this paper.)*

(iii) *Adaptive step lengths:*

$$\begin{aligned}\alpha_\nu := \max\Big\{\, \delta^j \;\Big|\; & f\big((1-\delta^j)u^\nu + \delta^j G(F(u^\nu))\big) - f(u^\nu) \\ & \le \big(f(u^\nu) - g(F(u^\nu))\big)(-\tfrac{1}{2}\delta^j),\ j\in\{0,1,2,\ldots\}\,\Big\},\\ \beta_\nu := \max\Big\{\, \delta^j \;\Big|\; & g(v^\nu) - g\big((1-\delta^j)v^\nu + \delta^j F(G(v^\nu))\big) \\ & \le \big(f(G(v^\nu)) - g(v^\nu)\big)(-\tfrac{1}{2}\delta^j),\ j\in\{0,1,2,\ldots\}\,\Big\}.\end{aligned}$$

Step 3 (update the iterates). *Use one of the following rules chosen at initialization to determine the next iterates.*

1. *Update with forward feedback:*

$$u^{\nu+1} := \operatorname{argmin}\{\, f(u) \,|\, u = \hat{u}^{\nu+1} \text{ or } u = G(\hat{v}^{\nu+1})\,\},$$
$$v^{\nu+1} := \operatorname{argmax}\{\, g(v) \,|\, v = \hat{v}^{\nu+1} \text{ or } v = F(\hat{u}^{\nu+1})\,\}.$$

(If both arguments give the same objective value, use the first one in updating for decisiveness. The same rule applies also to the next set of formulas.)

2. *Update with backward feedback:*

$$u^{\nu+1} := \operatorname{argmin}\{\, f(u) \,|\, u = \hat{u}^{\nu+1} \;\; \textit{or} \;\; u = G(v^\nu)\,\},$$
$$v^{\nu+1} := \operatorname{argmax}\{\, g(v) \,|\, v = \hat{v}^{\nu+1} \;\; \textit{or} \;\; v = F(u^\nu)\,\}.$$

Then return to Step 1 with the counter ν increased by 1.

Observe that the primal-dual feedback occurs in the updating of the iterates, as well as in the optimality test. This interaction links the primal and dual parts of the algorithm closely. For instance, with the optimality test in Step 1, the algorithm will terminate if either $u^\nu = \bar{u}$ or $v^\nu = \bar{v}$ by Theorem 1.1. The primal-dual feedback in Step 3 of updating has an important effect on the algorithm. In Sections 3 and 4, sharper estimates of the convergence ratio for the variants with the second update scheme will be obtained because of the backward feedback pattern in it. The reader is referred to [2] for more discussion on this.

In Step 2, there are three step length rules to choose from. The fixed step length in rule (ii) is probably of theoretical significance only, since $\tilde{\sigma}$ is unknown in most cases. However, Theorem 3.4 in the next section shows that the variant with the adaptive step length (iii) can achieve a global rate of convergence very close to the ones with step length (i) or (ii), provided that the parameter δ is chosen close to 1. The adaptive step length needs only some kind of backtracking computationally. When the problem is doubly decomposable, such as the box-separable case discussed in Section 5, the related computation can be massively parallelized, and even the perfect line search in rule (iii) will not be prohibitively difficult. In Section 4, we show that, in general, better asymptotic ratios than the global ones obtained in Section 3 can be expected for the variants with step rules (i) and (iii).

3. Global Linear Convergence of the PDSD Algorithm.

In this section, we derive global convergence results for all the six variants of the PDSD algorithm formulated in Section 2. Define the multivalued mappings $\mathcal{U}_0 : U \rightrightarrows U$, $\mathcal{V}_1 : U \rightrightarrows V$, $\mathcal{V}_0 : V \rightrightarrows V$ and $\mathcal{U}_1 : V \rightrightarrows U$ as follows.

$$\mathcal{U}_0(u) = [u, G(F(u))], \tag{3.1a}$$

$$\mathcal{V}_1(u) = \cup \{[F(u), F(\xi)] \mid \xi \in \mathcal{U}_0(u)\}, \tag{3.1b}$$

$$\mathcal{V}_0(v) = [v, F(G(v))], \tag{3.2a}$$

$$\mathcal{U}_1(v) = \cup \{[G(v), G(\eta)] \mid \eta \in \mathcal{V}_0(v)\}. \tag{3.2b}$$

We first give two lemmas that serve as bases for convergence results. Define the functions $\sigma_p : U \to \mathbb{R}$ and $\sigma_d : V \to \mathbb{R}$ as

$$\sigma_p(u) = \frac{M^2_{\mathcal{U}_0(u),F(u)}}{\lambda_{\mathcal{U}_0(u),F(u)}\lambda_{\mathcal{V}_1(u),\mathcal{U}_0(u)}} \text{ and } \sigma_d(v) = \frac{M^2_{G(v),\mathcal{V}_0(v)}}{\lambda_{\mathcal{V}_0(v),G(v)}\lambda_{\mathcal{U}_1(v),\mathcal{V}_0(v)}}, \tag{3.3}$$

where $M_{*,*}$ and $\lambda_{*,*}$ are defined in (2.1).

Lemma 3.1. *For any $u \in U$,*

$$f\big((1-\alpha)u + \alpha G(F(u))\big) - f(u) \le \big(f(u) - g(F(u))\big)\big(-\alpha + \sigma_p(u)\alpha^2\big) \tag{3.4}$$

for all $\alpha \in [0,1]$. Similarly, for any $v \in V$,

$$g(v) - g\big((1-\beta)v + \beta F(G(v))\big) \le \big(f(G(v)) - g(v)\big)\big(-\beta + \sigma_d(v)\beta^2\big) \tag{3.5}$$

for all $\beta \in [0,1]$.

Proof. For any $u_0 \in U$, denote $v_1 := F(u_0)$ and $u_2 := G(v_1) = G(F(u_0))$. Then by the mean value theorem [16], the Lagrangian $L(u,v)$ can be written in the expanded form at (u, v_1) as

$$L(u,v) = L(u,v_1) + \nabla_v L(u,v_1)\cdot(v - v_1) + \tfrac{1}{2}(v - v_1)\cdot\nabla^2_{vv} L(u,\tilde{v})(v - v_1)$$

with some $\tilde{v} \in [v_1, v]$ depending on u, v and v_1. Furthermore, the term $\nabla_v L(u,v_1)\cdot(v - v_1)$ can be written as

$$\nabla_v L(u,v_1)\cdot(v - v_1) = \nabla_v L(u_0,v_1)\cdot(v - v_1) + (v - v_1)\cdot\nabla^2_{vu} L(\tilde{u},v_1)(u - u_0)$$

with some $\tilde{u} \in [u_0, u]$. Note that $v_1 = F(u_0)$ means v_1 is the argmax of $L(u_0, v)$ on V, which in turn implies $\nabla_v L(u_0, v_1)\cdot(v - v_1) \le 0$ for all $v \in V$. Therefore

$$L(u,v) \le L(u,v_1) + (v-v_1)\cdot\nabla^2_{vu} L(\tilde{u}, v_1)(u-u_0) + \tfrac{1}{2}(v-v_1)\cdot\nabla^2_{vv} L(u,\tilde{v})(v-v_1). \quad (3.6)$$

Now for any $u \in [u_0, u_2]$ and $v = F(u)$, we have $\tilde{u} \in [u_0, u] \subset \mathcal{U}_0(u_0)$ and $\tilde{v} \in [v_1, F(u)] \subset \mathcal{V}_1(u_0)$. Hence it follows from (3.6) that

$$\begin{aligned}
&L\big(u, F(u)\big) - L(u, v_1) \\
&\le \big(F(u) - v_1\big)\cdot\nabla^2_{vu} L(\tilde{u}, v_1)(u - u_0) + \tfrac{1}{2}\big(F(u) - v_1\big)\cdot\nabla^2_{vv} L(u,\tilde{v})\big(F(u) - v_1\big) \\
&\le \max_{w \in \mathbb{R}^m} \{w\cdot\nabla^2_{vu} L(\tilde{u}, v_1)(u - u_0) + \tfrac{1}{2} w\cdot\nabla^2_{vv} L(u,\tilde{v}) w\} \\
&= \tfrac{1}{2}(u - u_0)\cdot\big(\nabla^2_{vu} L(\tilde{u}, v_1)\big)^T \big(-\nabla^2_{vv} L(u,\tilde{v})\big)^{-1} \nabla^2_{vu} L(\tilde{u}, v_1)(u - u_0) \\
&\le \tfrac{1}{2}\big(M^2_{\mathcal{U}_0(u_0),F(u_0)}/\lambda_{\mathcal{V}_1(u_0),\mathcal{U}_0(u_0)}\big)\|u - u_0\|^2 \qquad (3.7)
\end{aligned}$$

for all $u \in [u_0, u_2]$. Note that $L\big(u, F(u)\big) = f(u)$ and that

$$\begin{aligned}
L\big((1-\alpha)u_0 + \alpha u_2, v_1\big) &\le (1-\alpha)L(u_0, v_1) + \alpha L(u_2, v_1) \\
&= (1-\alpha) f(u_0) + \alpha g(v_1)
\end{aligned}$$

for $0 \le \alpha \le 1$. Thus, by taking $u = (1-\alpha)u_0 + \alpha u_2$ in (3.7), we get

$$\begin{aligned}
&f\big((1-\alpha)u_0 + \alpha u_2\big) - f(u_0) + \alpha\big(f(u_0) - g(v_1)\big) \\
&\le \tfrac{1}{2}\alpha^2\big(M^2_{\mathcal{U}_0(u_0),F(u_0)}/\lambda_{\mathcal{V}_1(u_0),\mathcal{U}_0(u_0)}\big)\|u_2 - u_0\|^2.
\end{aligned} \qquad (3.8)$$

Again by the mean value theorem,

$$\begin{aligned}
f(u_0) - g(v_1) &= L(u_0, v_1) - L(u_2, v_1) \\
&= \nabla_u L(u_2, v_1)\cdot(u_0 - u_2) + \tfrac{1}{2}(u_0 - u_2)\cdot\nabla^2_{uu} L(u', v_1)(u_0 - u_2)
\end{aligned}$$

for some $u' \in [u_0, u_2] = \mathcal{U}_0(u_0)$. Observe that $\nabla_u L(u_2, v_1)\cdot(u_0 - u_2) \ge 0$, since u_2 is the argmin of $L(u, v_1)$ on U. Therefore

$$\begin{aligned}
f(u_0) - g(v_1) &\ge \tfrac{1}{2}(u_0 - u_2)\cdot\nabla^2_{uu} L(u', v_1)(u_0 - u_2) \\
&\ge \tfrac{1}{2}\lambda_{\mathcal{U}_0(u),F(u)}\|u_2 - u_0\|^2
\end{aligned}$$

or, equivalently,

$$\big(f(u_0) - g(v_1)\big)/\lambda_{\mathcal{U}_0(u_0),F(u_0)} \ge \tfrac{1}{2}\|u_2 - u_0\|^2. \qquad (3.9)$$

Combining (3.8) and (3.9), we get

$$
\begin{aligned}
f\big((1-\alpha)u_0+\alpha u_2\big)-f(u_0) &\le \big(f(u_0)-g(v_1)\big)\big(-\alpha+\frac{\alpha^2 M^2_{\mathcal{U}_0(u_0),F(u_0)}}{\lambda_{\mathcal{U}_0(u_0),F(u_0)}\lambda_{\mathcal{V}_1(u_0),\mathcal{U}_0(u_0)}}\big)\\
&\le \big(f(u_0)-g(v_1)\big)\big(-\alpha+\sigma_p(u_0)\alpha^2\big) \qquad (3.10)
\end{aligned}
$$

for $0\le\alpha\le 1$, which yields (3.4). We can prove (3.5) similarly. $\square$

Lemma 3.2. *Suppose $u^\nu\in U$, $v^\nu\in V$, $\alpha_\nu\in[0,1]$ and $\beta_\nu\in[0,1]$ for some integer $\nu\ge 0$. If there exists $\zeta_\nu>0$ such that*

$$
f\big((1-\alpha_\nu)u^\nu+\alpha_\nu G(F(u^\nu))\big)-f(u^\nu)\le\big(f(u^\nu)-g(F(u^\nu))\big)(-\zeta_\nu), \qquad (3.11)
$$

$$
g(v^\nu)-g\big((1-\beta_\nu)v^\nu+\beta_\nu F(G(v^\nu))\big)\le\big(f(G(v^\nu))-g(v^\nu)\big)(-\zeta_\nu), \qquad (3.12)
$$

then in the PDSD algorithm,

(a) *the new pair $(u^{\nu+1},v^{\nu+1})$ generated by the update scheme 1 satisfies*

$$
f(u^{\nu+1})-f(\bar u)\le(1-\zeta_\nu)\big(f(u^\nu)-f(\bar u)\big), \qquad (3.13)
$$

$$
g(\bar v)-g(v^{\nu+1})\le(1-\zeta_\nu)\big(g(\bar v)-g(v^\nu)\big), \qquad (3.14)
$$

(b) *while the new pair $(u^{\nu+1},v^{\nu+1})$ generated by the update scheme 2 satisfies*

$$
f(u^{\nu+1})-g(v^{\nu+1})\le\frac{1-\zeta_\nu}{1+\zeta_\nu}\big(f(u^\nu)-g(v^\nu)\big). \qquad (3.15)
$$

Proof. According to the update scheme in Step 3 of the algorithm, we have

$$
\begin{aligned}
f(u^{\nu+1})&\le f\big((1-\alpha_\nu)u^\nu+\alpha_\nu G(F(u^\nu))\big),\\
g(v^{\nu+1})&\ge g\big((1-\beta_\nu)v^\nu+\beta_\nu F(G(v^\nu))\big).
\end{aligned}
$$

Hence (3.11) and (3.12) imply

$$
f(u^{\nu+1})-f(u^\nu)\le\big(f(u^\nu)-g(F(u^\nu))\big)(-\zeta_\nu), \qquad (3.16)
$$

$$
g(v^\nu)-g(v^{\nu+1})\le\big(f(G(v^\nu))-g(v^\nu)\big)(-\zeta_\nu). \qquad (3.17)
$$

Define the νth *duality gap* ε_ν and the νth *auxiliary duality gap* $\tilde\varepsilon_\nu$ as

$$
\varepsilon_\nu:=f(u^\nu)-g(v^\nu)\ \text{and}\ \tilde\varepsilon_\nu:=f(G(v^\nu))-g(F(u^\nu)) \qquad (3.18)
$$

respectively. By combining (3.16) and (3.17), we get

$$\varepsilon_{\nu+1} \le (1-\zeta_\nu)\varepsilon_\nu - \zeta_\nu \tilde{\varepsilon}_\nu. \tag{3.19}$$

If update scheme 2 is used, then $f(u^{\nu+1}) \le f(G(v^\nu))$ and $g(v^{\nu+1}) \ge g(F(u^\nu))$. Hence $\varepsilon_{\nu+1} \le \tilde{\varepsilon}_\nu$. Therefore (3.19) implies

$$\varepsilon_{\nu+1} \le (1-\zeta_\nu)\varepsilon_\nu - \zeta_\nu \varepsilon_{\nu+1},$$

which yields (3.15).

On the other hand, if update scheme 1 is used, then the relation $\varepsilon_{\nu+1} \le \tilde{\varepsilon}_\nu$ is not necessarily ture. However, by Theorem 1.1,

$$f(u) \ge f(\bar{u}) = g(\bar{v}) \ge g(v) \qquad \text{for all } u \in U,\ v \in V.$$

Hence if follows from (3.16) and (3.17) that

$$\begin{aligned} f(u^{\nu+1}) - f(u^\nu) &\le \big(f(u^\nu) - f(\bar{u})\big)(-\zeta_\nu), \\ g(v^\nu) - g(v^{\nu+1}) &\le \big(g(\bar{v}) - g(v^\nu)\big)(-\zeta_\nu). \end{aligned}$$

These two inequalities yield (3.13) and (3.14) respectively. □

Define the function $\theta : [0,+\infty) \to (0,1)$ as

$$\theta(s) = \begin{cases} 1-s & \text{if } s < \frac{1}{2}, \\ \frac{1}{4s} & \text{if } s \ge \frac{1}{2}. \end{cases} \tag{3.20}$$

The following theorem gives convergence results for the algorithms with perfect line search (i) and fixed step length (ii).

Theorem 3.3 (convergence of PDSD with step length rules (i) and (ii)).

(a) *The sequences $\{f(u^\nu)\}$ and $\{g(v^\nu)\}$ generated by PDSD–1(i) or PDSD–1(ii) converge linearly to the common optimal value $f(\bar{u}) = g(\bar{v})$ in the sense that*

$$\begin{aligned} f(u^{\nu+1}) - f(\bar{u}) &\le \big(1-\theta(\tilde{\sigma})\big)\big(f(u^\nu) - f(\bar{u})\big) \\ &= \begin{cases} \tilde{\sigma}\big(f(u^\nu) - f(\bar{u})\big) & \textit{if } 0 \le \tilde{\sigma} < \frac{1}{2}, \\ \big(1 - 1/(4\tilde{\sigma})\big)\big(f(u^\nu) - f(\bar{u})\big) & \textit{if } \tilde{\sigma} \ge \frac{1}{2}, \end{cases} \end{aligned} \tag{3.21}$$

$$\begin{aligned} g(\bar{v}) - g(v^{\nu+1}) &\le \big(1-\theta(\tilde{\sigma})\big)\big(g(\bar{v}) - g(v^\nu)\big) \\ &= \begin{cases} \tilde{\sigma}\big(g(\bar{v}) - g(v^\nu)\big) & \textit{if } 0 \le \tilde{\sigma} < \frac{1}{2}, \\ \big(1 - 1/(4\tilde{\sigma})\big)\big(g(\bar{v}) - g(v^\nu)\big) & \textit{if } \tilde{\sigma} \ge \frac{1}{2}. \end{cases} \end{aligned} \tag{3.22}$$

Moreover,

$$\lambda_{U,V}\|u^{\nu+1}-\bar u\|^2+\lambda_{V,U}\|v^{\nu+1}-\bar v\|^2 \le 2\big(1-\theta(\tilde\sigma)\big)^{\nu+1}\big(f(u^0)-g(u^0)\big). \tag{3.23}$$

(b) *The sequences* $\{f(u^\nu)\}$ *and* $\{g(v^\nu)\}$ *generated by PDSD–2(i) or PDSD–2(ii) converge linearly to the common optimal value* $f(\bar u)=g(\bar v)$ *in the sense that*

$$\begin{aligned} f(u^{\nu+1})-g(v^{\nu+1}) &\le \frac{1-\theta(\tilde\sigma)}{1+\theta(\tilde\sigma)}\big(f(u^\nu)-g(v^\nu)\big),\\ &= \begin{cases} \big(\tilde\sigma/(2-\tilde\sigma)\big)\big(f(u^\nu)-g(v^\nu)\big) & \text{if } 0\le\tilde\sigma<\frac12,\\ \big(1-1/(2\tilde\sigma+0.5)\big)\big(f(u^\nu)-g(v^\nu)\big) & \text{if } \tilde\sigma\ge\frac12.\end{cases}\end{aligned} \tag{3.24}$$

Moreover,

$$\lambda_{U,V}\|u^{\nu+1}-\bar u\|^2+\lambda_{V,U}\|v^{\nu+1}-\bar v\|^2 \le 2\Big(\frac{1-\theta(\tilde\sigma)}{1+\theta(\tilde\sigma)}\Big)^{\nu+1}\big(f(u^0)-g(u^0)\big). \tag{3.25}$$

Proof. Note that $\mathcal{U}_0(u)$ and $\mathcal{V}_1(u)$ are subsets of U, while $\mathcal{V}_0(v)$ and $\mathcal{U}_1(v)$ are subsets of V. Hence for any $u\in U$ and $v\in V$,

$$\begin{aligned} &\lambda_{\mathcal{U}_0(u),F(u)}\ge\lambda_{U,V},\ \lambda_{\mathcal{V}_1(u),\mathcal{U}_0(u)}\ge\lambda_{V,U},\ M_{\mathcal{U}_0(u),F(u)}\le M_{U,V},\\ &\lambda_{\mathcal{V}_0(v),G(v)}\ge\lambda_{V,U},\ \lambda_{\mathcal{U}_1(v),\mathcal{V}_0(v)}\ge\lambda_{U,V},\ M_{G(v),\mathcal{V}_0(v)}\le M_{U,V}.\end{aligned}$$

Therefore,

$$\tilde\sigma\ge\sigma_p(u) \quad\text{and}\quad \tilde\sigma\ge\sigma_d(v) \tag{3.26}$$

for all $u\in U$ and $v\in V$. Substituting (3.26) in (3.4) and (3.5), we get

$$f\big((1-\alpha)u+\alpha G(F(u))\big)-f(u)\le\big(f(u)-g(F(u))\big)(-\alpha+\tilde\sigma\alpha^2), \tag{3.27}$$

$$g(v)-g\big((1-\beta)v+\beta F(G(v))\big)\le\big(f(G(v))-g(v)\big)\big(-\beta+\tilde\sigma\beta^2\big), \tag{3.28}$$

for all $\alpha\in[0,1]$ and $\beta\in[0,1]$.

Observe that $\min\{-\alpha+\tilde\sigma\alpha^2 \mid 0\le\alpha\le1\}=-\theta(\tilde\sigma)<0$ with

$$\operatorname{argmin}\{-\alpha+\tilde\sigma\alpha^2 \mid 0\le\alpha\le1\}=\min\{1,\frac{1}{2\tilde\sigma}\}.$$

Hence for the fixed step length $\alpha_\nu=\min\{1,\frac{1}{2\tilde\sigma}\}$ in rule (ii), (3.27) implies

$$f\big((1-\alpha_\nu)u^\nu+\alpha_\nu G(F(u^\nu))\big)-f(u^\nu)\le\big(f(u^\nu)-g(F(u^\nu))\big)\big(-\theta(\tilde\sigma)\big). \tag{3.29}$$

Obviously (3.29) is also true for the step length $\alpha_\nu = \operatorname*{argmin}_{\alpha\in[0,1]} f\big((1-\alpha)u^\nu+\alpha G(F(u^\nu))\big)$ in rule (i), since the perfect line search should not make the first term of (3.29) any larger. According to the update scheme in Step 3, we have $f(u^{\nu+1}) \le f((1-\alpha_\nu)u^\nu + \alpha_\nu G(F(u^\nu)))$. Therefore

$$f(u^{\nu+1}) - f(u^\nu) \le \big(f(u^\nu) - g(F(u^\nu))\big)\big(-\theta(\tilde{\sigma})\big). \tag{3.30}$$

Similarly it follows from (3.28) that in the dual variable,

$$g(v^\nu) - g(v^{\nu+1}) \le \big(f(G(v^\nu)) - g(v^\nu)\big)\big(-\theta(\tilde{\sigma})\big). \tag{3.31}$$

Now by invoking Lemma 3.2 with $\zeta_\nu = \theta(\tilde{\sigma})$, we get (3.21), (3.22) and (3.24). Using (3.24) for $\nu = 0, 1, \ldots$, we get

$$f(u^{\nu+1}) - g(v^{\nu+1}) \le \Big(\frac{1-\theta(\tilde{\sigma})}{1+\theta(\tilde{\sigma})}\Big)^{\nu+1}\big(f(u^0) - g(v^0)\big),$$

which yields (3.25) for PDSD–2 by Proposition 2.2.

Note that $f(\bar{u}) = g(\bar{v})$. Combining (3.21) and (3.22), we get

$$f(u^{\nu+1}) - g(v^{\nu+1}) \le \big(1-\theta(\tilde{\sigma})\big)\big(f(u^\nu) - g(v^\nu)\big). \tag{3.32}$$

Using (3.32) for $\nu = 0, 1, \ldots$, we get

$$f(u^{\nu+1}) - g(v^{\nu+1}) \le \big(1-\theta(\tilde{\sigma})\big)^{\nu+1}\big(f(u^0) - g(v^0)\big),$$

which yields (3.23) for PDSD–1 by Proposition 2.2. $\square$

Next we give convergence results for the algorithm with adaptive step lengths (iii). We have to show, in the first place, that these step lengths are well defined. Let the function $\tilde{\theta} : [0, +\infty) \to (0, 1)$ be defined as

$$\tilde{\theta}(s) = \min\{\frac{1}{2}, \frac{1}{4s}\}. \tag{3.33}$$

Obviously $\theta(s) \ge \tilde{\theta}(s)$ for all $s \in [0, +\infty)$, and the equality holds when $s \ge \frac{1}{2}$.

Theorem 3.4 (convergence of PDSD with step length rule (iii)).

(a) *For any $\delta \in (0,1)$, the step lengths α_ν and β_ν in the PDSD algorithm with rule (iii) are well defined. Moreover,*

$$\alpha_\nu > \delta \min\{1, \frac{1}{2\tilde{\sigma}}\} \quad \text{and} \quad \beta_\nu > \delta \min\{1, \frac{1}{2\tilde{\sigma}}\} \tag{3.34}$$

for all ν.

(b) *The sequences $\{f(u^\nu)\}$ and $\{g(v^\nu)\}$ generated by PDSD–1(iii) converge linearly to the common optimal value $f(\bar{u}) = g(\bar{v})$ in the sense that*

$$\begin{aligned} f(u^{\nu+1}) - f(\bar{u}) &\le \big(1 - \delta\tilde{\theta}(\tilde{\sigma})\big)\big(f(u^\nu) - f(\bar{u})\big) \\ &= \begin{cases} (\delta/2)\big(f(u^\nu) - f(\bar{u})\big) & \text{if } 0 \le \tilde{\sigma} < \frac{1}{2}, \\ \big(1 - \delta/(4\tilde{\sigma})\big)\big(f(u^\nu) - f(\bar{u})\big) & \text{if } \tilde{\sigma} \ge \frac{1}{2}, \end{cases} \end{aligned} \tag{3.35}$$

$$\begin{aligned} g(\bar{v}) - g(v^{\nu+1}) &\le \big(1 - \delta\tilde{\theta}(\tilde{\sigma})\big)\big(g(\bar{v}) - g(v^\nu)\big) \\ &= \begin{cases} (\delta/2)\big(g(\bar{v}) - g(v^\nu)\big) & \text{if } 0 \le \tilde{\sigma} < \frac{1}{2}, \\ \big(1 - \delta/(4\tilde{\sigma})\big)\big(g(\bar{v}) - g(v^\nu)\big) & \text{if } \tilde{\sigma} \ge \frac{1}{2}. \end{cases} \end{aligned} \tag{3.36}$$

Moreover

$$\lambda_{U,V}\|u^{\nu+1} - \bar{u}\| + \lambda_{V,U}\|v^{\nu+1} - \bar{v}\| \le 2\big(1 - \delta\tilde{\theta}(\tilde{\sigma})\big)^{\nu+1}\big(f(u^0) - g(u^0)\big). \tag{3.37}$$

(c) *The sequences $\{f(u^\nu)\}$ and $\{g(v^\nu)\}$ generated by PDSD–2(iii) converge linearly to the common optimal value $f(\bar{u}) = g(\bar{v})$ in the sense that*

$$\begin{aligned} f(u^{\nu+1}) - g(v^{\nu+1}) &\le \frac{1 - \delta\tilde{\theta}(\tilde{\sigma})}{1 + \delta\tilde{\theta}(\tilde{\sigma})}\big(f(u^\nu) - g(v^\nu)\big), \\ &= \begin{cases} \big((2-\delta)/(2+\delta)\big)\big(f(u^\nu) - g(v^\nu)\big) & \text{if } 0 \le \tilde{\sigma} < \frac{1}{2}, \\ \big(1 - 1/(2\delta\tilde{\sigma} + 0.5)\big)\big(f(u^\nu) - g(v^\nu)\big) & \text{if } \tilde{\sigma} \ge \frac{1}{2}. \end{cases} \end{aligned} \tag{3.38}$$

Moreover

$$\lambda_{U,V}\|u^{\nu+1} - \bar{u}\|^2 + \lambda_{V,U}\|v^{\nu+1} - \bar{v}\|^2 \le 2\Big(\frac{1 - \delta\tilde{\theta}(\tilde{\sigma})}{1 + \delta\tilde{\theta}(\tilde{\sigma})}\Big)^{\nu+1}\big(f(u^0) - g(u^0)\big). \tag{3.39}$$

Proof. First, we claim that for all nonnegative $\alpha \le \min\{1, \frac{1}{2\sigma_p(u^\nu)}\}$,

$$f\big((1-\alpha)u^\nu + \alpha G(F(u^\nu))\big) - f(u^\nu) \le \big(f(u^\nu) - g(F(u^\nu))\big)\big(\frac{-\alpha}{2}\big). \tag{3.40}$$

This follows directly from (3.4) and the fact that

$$-\alpha + \sigma_p(u^\nu)\alpha^2 \le \frac{-\alpha}{2} \quad \text{for all } 0 \le \alpha \le \min\{1, \frac{1}{2\sigma_p(u^\nu)}\}.$$

Hence the step length $\alpha_\nu = \delta^j$ in rule (iii), where j is the *first* element in the ordered nonnegative integer set $\{0, 1, 2, \ldots\}$ satisfying

$$f\big((1-\delta^j)u^\nu + \delta^j G(F(u^\nu))\big) - f(u^\nu) \le \big(f(u^\nu) - g(F(u^\nu))\big)(-\tfrac{1}{2}\delta^j), \tag{3.41}$$

is well defined. Moreover, $\delta^{j-1} > \min\{1, \frac{1}{2\sigma_p(u^\nu)}\}$ if $j \ne 0$, because otherwise δ^{j-1} instead of δ^j will be taken as the step length according to rule (iii). Thus

$$\alpha_\nu = \delta^j > \delta \min\{1, \frac{1}{2\sigma_p(u^\nu)}\} = 2\delta\tilde{\theta}(\sigma_p(u^\nu)). \tag{3.42}$$

Combining (3.41) and (3.42), we have

$$\begin{aligned} f\big((1-\alpha_\nu)u^\nu + \alpha_\nu G(F(u^\nu))\big) - f(u^\nu) &\le \big(f(u^\nu) - g(F(u^\nu))\big)\big(-\delta\tilde{\theta}(\sigma_p(u^\nu))\big), \\ &\le \big(f(u^\nu) - g(F(u^\nu))\big)\big(-\delta\tilde{\theta}(\tilde{\sigma})\big), \end{aligned} \tag{3.43}$$

where the last inequality follows from (3.26) and (3.33). Similarly,

$$\begin{aligned} g(v^\nu) - g\big((1-\beta_\nu)v^\nu + \beta_\nu F(G(v^\nu))\big) &\le \big(f(G(v^\nu)) - g(v^\nu)\big)\big(-\delta\tilde{\theta}(\sigma_d(u^\nu))\big), \\ &\le \big(f(G(v^\nu)) - g(v^\nu)\big)\big(-\delta\tilde{\theta}(\tilde{\sigma})\big). \end{aligned} \tag{3.44}$$

Now by invoking Lemma 3.2 with $\zeta_\nu = \delta\tilde{\theta}(\tilde{\sigma})$, we get (3.35), (3.36) and (3.38). The inequality (3.37) follows from (3.35) and (3.36) in the same manner as (3.23) follows from (3.21) and (3.22) in Theorem 3.3. Similarly, (3.39) follows from (3.38) in the same manner as (3.25) follows from (3.24) in Theorem 3.3. □

4. Asymptotic Rates of Convergence.

Define

$$\bar{\sigma} = \frac{M^2_{\bar{u},\bar{v}}}{\lambda_{\bar{u},\bar{v}}\lambda_{\bar{v},\bar{u}}}. \tag{4.1}$$

Note that $\bar{\sigma}$ is related only to quantities defined at the optimal solution. Obviously, we have $\lambda_{\bar{u},\bar{v}} \ge \lambda_{U,V}$, $\lambda_{\bar{v},\bar{u}} \ge \lambda_{V,U}$ and $M_{\bar{v},\bar{u}} \le M_{V,U}$. Therefore $\bar{\sigma} \le \tilde{\sigma}$. In this section, we show that it is this smaller number $\bar{\sigma}$ that governs the rate of convergence in the tail of iteration for the variants with perfect line search or adaptive step length.

Theorem 4.1 (asymptotic rates of PDSD with step length rule (i)). *Suppose the algorithm does not terminate at the solution pair $(\bar{u}, \bar{v})$ after a finite number of iterations, then*

(a) *for the sequences $\{f(u^\nu)\}$ and $\{g(v^\nu)\}$ generated by PDSD–1(i),*

$$\limsup_{\nu\to\infty} \frac{f(u^{\nu+1}) - f(\bar{u})}{f(u^\nu) - f(\bar{u})} \le 1 - \theta(\bar{\sigma}) = \begin{cases} \bar{\sigma} & \text{if } 0 \le \bar{\sigma} < \frac{1}{2}, \\ 1 - 1/(4\bar{\sigma}) & \text{if } \bar{\sigma} \ge \frac{1}{2}, \end{cases} \tag{4.2}$$

$$\limsup_{\nu\to\infty} \frac{g(\bar{v}) - g(v^{\nu+1})}{g(\bar{v}) - g(v^\nu)} \le 1 - \theta(\bar{\sigma}) = \begin{cases} \bar{\sigma} & \text{if } 0 \le \bar{\sigma} < \frac{1}{2}, \\ 1 - 1/(4\bar{\sigma}) & \text{if } \bar{\sigma} \ge \frac{1}{2}, \end{cases} \tag{4.3}$$

(b) while for the sequences $\{f(u^\nu)\}$ and $\{g(v^\nu)\}$ generated by PDSD–2(i),

$$\limsup_{\nu\to\infty} \frac{f(u^{\nu+1}) - g(v^{\nu+1})}{f(u^\nu) - g(v^\nu)} \le \frac{1 - \theta(\bar{\sigma})}{1 + \theta(\bar{\sigma})} = \begin{cases} \bar{\sigma}/(2 - \bar{\sigma}) & \text{if } 0 \le \bar{\sigma} < \frac{1}{2}, \\ 1 - 1/(2\bar{\sigma} + 0.5) & \text{if } \bar{\sigma} \ge \frac{1}{2}. \end{cases} \tag{4.4}$$

Proof. Observe that the algorithm will terminate at $(\bar{u}, \bar{v})$ if either $u^\nu = \bar{u}$ or $v^\nu = \bar{v}$. Hence under the assumption of the theorem, the denominators in (4.2)–(4.3) are positive. By Lemma 3.2, we need only to prove that for any positive $\zeta < \theta(\bar{\sigma})$, there exists an integer $\bar{\nu}$ such that for all $\nu \ge \bar{\nu}$,

$$f\big((1 - \alpha_\nu)u^\nu + \alpha_\nu G(F(u^\nu))\big) - f(u^\nu) \le \big(f(u^\nu) - g(F(u^\nu))\big)(-\zeta), \tag{4.5}$$

$$g(v^\nu) - g\big((1 - \beta_\nu)v^\nu + \beta_\nu F(G(v^\nu))\big) \le \big(f(G(v^\nu)) - g(v^\nu)\big)(-\zeta). \tag{4.6}$$

Observe that $\min\{\, -\alpha + \sigma_p(u^\nu)\alpha^2 \mid 0 \le \alpha \le 1 \,\} = -\theta(\sigma_p(u^\nu)) < 0$, with

$$\operatorname{argmin}\{\, -\alpha + \sigma_p(u^\nu)\alpha^2 \mid 0 \le \alpha \le 1 \,\} = \min\{1, \frac{1}{2\sigma_p(u^\nu)}\}.$$

Hence, it follows from (3.4) in Lemma 3.1 that

$$f\big((1 - \alpha_\nu)u^\nu + \alpha_\nu G(F(u^\nu))\big) - f(u^\nu) \le \big(f(u^\nu) - g(F(u^\nu))\big)\big(-\theta(\sigma_p(u^\nu))\big). \tag{4.7}$$

Similarly,

$$g(v^\nu) - g\big((1 - \beta_\nu)v^\nu + \beta_\nu F(G(v^\nu))\big) \le \big(f(G(v^\nu)) - g(v^\nu)\big)\big(-\theta(\sigma_d(v^\nu))\big). \tag{4.8}$$

Now according to Theorem 4.3, $u^\nu \to \bar{u}$ and $v^\nu \to \bar{v}$ as $\nu \to \infty$. Recall that $\bar{v} = F(\bar{u})$ and $\bar{u} = G(\bar{v})$ by Theorem 1.1. Hence it follows from Proposition 2.3 that

$$G(v^\nu) \to \bar{u},\ F(u^\nu) \to \bar{v},\ G(F(u^\nu)) \to \bar{u},\ F(G(v^\nu)) \to \bar{v} \text{ as } \nu \to \infty.$$

Then for the sets defined in (3.1) and (3.2),

$$\mathrm{dist}(\mathcal{U}_0(u^\nu),\bar u)\to 0,\ \mathrm{dist}(\mathcal{V}_1(u^\nu),\bar u)\to 0,\ \mathrm{dist}(\mathcal{V}_0(v^\nu),\bar v)\to 0,\ \mathrm{dist}(\mathcal{U}_1(v^\nu),\bar v)\to 0,$$

as $\nu\to\infty$, which implies

$$M_{\mathcal{U}_0(u^\nu),F(u^\nu)}\to M_{\bar u,\bar v},\ M_{\mathcal{V}_0(v^\nu),G(v^\nu)}\to M_{\bar v,\bar u},\ \lambda_{\mathcal{U}_1(v^\nu),\mathcal{V}_0(v^\nu)}\to\lambda_{\bar u,\bar v},$$
$$\lambda_{\mathcal{V}_1(u^\nu),\mathcal{U}_0(u^\nu)}\to\lambda_{\bar v,\bar u},\ \lambda_{\mathcal{U}_0(u^\nu),F(u^\nu)}\to\lambda_{\bar u,\bar v},\ \lambda_{\mathcal{V}_0(v^\nu),G(v^\nu)}\to\lambda_{\bar v,\bar u},$$

since the Lagrangian $L(u,v)$ is C^2. Therefore

$$\sigma_p(u^\nu)\to\bar\sigma \text{ and } \sigma_d(v^\nu)\to\bar\sigma$$

as $\nu\to\infty$. Hence

$$\theta(\sigma_p(u^\nu))\to\theta(\bar\sigma) \text{ and } \theta(\sigma_d(v^\nu))\to\theta(\bar\sigma)$$

as $\nu\to\infty$. Thus the claim is true by (4.7) and (4.8). This completes the proof. □

Theorem 4.2 (asymptotic rates of PDSD with step length rule (iii))**.** *Suppose the algorithm does not terminate at the solution pair $(\bar u,\bar v)$ after a finite number of iterations, then*

(a) *for the sequences $\{f(u^\nu)\}$ and $\{g(v^\nu)\}$ generated by PDSD–1(iii),*

$$\limsup_{\nu\to\infty}\frac{f(u^{\nu+1})-f(\bar u)}{f(u^\nu)-f(\bar u)}\le 1-\delta\tilde\theta(\bar\sigma)=\begin{cases}\delta/2 & \text{if } 0\le\bar\sigma<\frac12,\\ 1-\delta/(4\bar\sigma) & \text{if } \bar\sigma\ge\frac12,\end{cases}\tag{4.9}$$
$$\limsup_{\nu\to\infty}\frac{g(\bar v)-g(v^{\nu+1})}{g(\bar v)-g(v^\nu)}\le 1-\delta\theta(\bar\sigma)=\begin{cases}\delta/2 & \text{if } 0\le\bar\sigma<\frac12,\\ 1-\delta/(4\bar\sigma) & \text{if } \bar\sigma\ge\frac12,\end{cases}\tag{4.10}$$

(b) while for the sequences $\{f(u^\nu)\}$ and $\{g(v^\nu)\}$ generated by PDSD–2(i),

$$\limsup_{\nu\to\infty}\frac{f(u^{\nu+1})-g(v^{\nu+1})}{f(u^\nu)-g(v^\nu)}\le\frac{1-\delta\theta(\bar\sigma)}{1+\delta\theta(\bar\sigma)}=\begin{cases}(2-\delta)/(2+\delta) & \text{if } 0\le\bar\sigma<\frac12,\\ 1-1/(2\delta\bar\sigma+0.5) & \text{if } \bar\sigma\ge\frac12.\end{cases}\tag{4.11}$$

Proof. Note that in the proof of Theorem 3.4, we have proved

$$f\big((1-\alpha_\nu)u^\nu+\alpha_\nu G(F(u^\nu))\big)-f(u^\nu)\le\big(f(u^\nu)-g(F(u^\nu))\big)\big(-\delta\tilde\theta(\sigma_p(u^\nu))\big),$$
$$g(v^\nu)-g\big((1-\beta_\nu)v^\nu+\beta_\nu F(G(v^\nu))\big)\le\big(f(G(v^\nu))-g(v^\nu)\big)\big(-\delta\tilde\theta(\sigma_d(u^\nu))\big),$$

as the first halves of (3.43) and (3.44). These two inequalities lead to the conclusions of the theorem in the same manner as (4.7) and (4.8) lead to the conclusions of Theorem 4.1. □

5. Minimax Problems with LCT.

As an example of a potential large-scale application of the algorithm, we discuss the minimax problem on $U \times V$ with the Lagrangian

$$L(u, v) = \varphi(u) - \psi(v) - v \cdot Ru, \tag{5.1}$$

where the matrix R is in $\mathbb{R}^{m \times n}$, and the functions $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $\psi : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are closed proper convex with their effective domains satisfying

$$U \subset \operatorname{dom} \varphi \text{ and } V \subset \operatorname{dom} \psi. \tag{5.2}$$

We refer to such a problem as *the minimax problem with linear cross terms* (with LCT for short). The inclusion (5.2) means that there are no implicit constraints other than the explicit one $(u, v) \in U \times V$ for the minimax problem.

Define the functions

$$\begin{aligned} \rho_{V,\psi}(r) &= \sup_{v \in V} \{ r \cdot v - \psi(v) \} \text{ for } r \in \mathbb{R}^m, \\ \rho_{U,\varphi}(s) &= \sup_{u \in U} \{ s \cdot u - \varphi(u) \} \text{ for } s \in \mathbb{R}^n. \end{aligned} \tag{5.3}$$

The objective functions in $(\mathcal{P})$ and $(\mathcal{Q})$ can be written as

$$f(u) = \varphi(u) + \rho_{V,\psi}(-Ru) \quad \text{and} \quad g(v) = -\psi(v) - \rho_{U,\varphi}(R^T v) \tag{5.4}$$

(where the "T" signals the transpose matrix). If U and V are polyhedral convex sets, and φ and ψ are linear-quadratic convex functions, then problems $(\mathcal{P})$ and $(\mathcal{Q})$ reduce to the ELQP discussed in [3–11].

As an instance fitting the concept of double decomposability, consider the following *box-separable* case, where the functions $\varphi(u)$ and $\psi(v)$ are separable

$$\varphi(u) = \sum_{j=1}^{n} \varphi_j(u_j), \quad \psi(v) = \sum_{i=1}^{m} \psi_i(v_i),$$

and U and V are Cartesian products of intervals (not necessarily finite):

$$U = [u_1^-, u_1^+] \times \ldots \times [u_n^-, u_n^+], \quad V = [v_1^-, v_1^+] \times \ldots \times [v_m^-, v_m^+].$$

The primal problem $(\mathcal{P})$ then takes the form of minimizing

$$f(u) = \sum_{j=1}^{n} \varphi_j(u_j) + \sum_{i=1}^{m} \rho_{v_i^-, v_i^+, \psi_i}\left(-\sum_{j=1}^{n} r_{ij} u_j\right) \tag{5.5}$$

subject to $u_j^- \le u_j \le u_j^+$ for $j = 1, \ldots, n$, where

$$\rho_{v_i^-, v_i^+, \psi_i}(w_i) = \sup_{v_i^- \le v_i \le v_i^+} \{w_i v_i - \psi_i(v_i)\}. \tag{5.6}$$

Hence the maximization of the Lagrangian in the calculation of $f(u)$ and $F(u)$ can be decomposed to a set of one-dimensional problems of the type in (5.6)(similarly for the calculations of $g(v)$ and $G(v)$.) Therefore the problem is doubly decomposable, and the computations related to the mappings F and G in the algorithm can be massively parallelized.

The ρ terms (or the *monitoring function* in the terminology of [8, 9]) in ELQP can represent penalties of piecewise linear-quadratic nature, as well as sharp linear inequality or equality constraints [6]. Now with φ and ψ being more general convex functions than the linear-quadratic ones, the formulation will provide even richer possibilities. The following proposition points out that the ρ terms in the box-separable case are the corresponding conjugate functions [12] extrapolated to the left and right by linear functions.

Proposition 5.1 (ρ terms in the box-separable case). *Suppose $v_i^+ > v_i^-$ and $\psi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is a closed proper convex function with $]v_i^-, v_i^+[\subset \operatorname{dom} \psi_i$, where $\operatorname{dom} \psi_i$ is the effective domain of ψ_i. Let ψ_i^* be the conjugate of ψ_i*

$$\psi_i^*(w_i) = \sup_{v_i \in \mathbb{R}} \{w_i v_i - \psi_i(v_i)\}.$$

Define

$$\begin{aligned} w_i^+ &= \begin{cases} \sup\{w_i \in \mathbb{R} \mid [v_i^-, v_i^+] \cap \partial \psi_i^*(w_i) \neq \emptyset\} & \text{if } v_i^+ \neq +\infty, \\ +\infty & \text{if } v_i^+ = +\infty, \end{cases} \\ w_i^- &= \begin{cases} \inf\{w_i \in \mathbb{R} \mid [v_i^-, v_i^+] \cap \partial \psi_i^*(w_i) \neq \emptyset\} & \text{if } v_i^- \neq -\infty, \\ -\infty & \text{if } v_i^- = -\infty. \end{cases} \end{aligned} \tag{5.7}$$

Then

$$\rho_{v_i^-, v_i^+, \psi_i}(w_i) = \begin{cases} v_i^+(w_i - w_i^+) + \psi_i^*(w_i^+) & \text{if } w_i^+ \le w_i, \\ \psi_i^*(w_i) & \text{if } w_i^- < w_i < w_i^+, \\ v_i^-(w_i - w_i^-) + \psi_i^*(w_i^-) & \text{if } w_i \le w_i^-. \end{cases} \tag{5.8}$$

Proof. We claim that the set

$$W_i = \{w_i \in \mathbb{R} \mid [v_i^-, v_i^+] \cap \partial\psi_i^*(w_i) \neq \emptyset\}$$

is nonempty and is an interval. To prove this, we observe that $]v_i^-, v_i^+[\, \cap \, \mathrm{ri}(\mathrm{dom}\, \psi_i)$ is nonempty, since $v_i^+ > v_i^-$ and $]v_i^-, v_i^+[\, \subset \mathrm{dom}\, \psi_i$. Hence there exists

$$\hat{v}_i \in \,]v_i^-, v_i^+[\text{ such that } \partial\psi_i(\hat{v}_i) \neq \emptyset \tag{5.9}$$

by [12, Theorem 23.4]. Then

$$\hat{v}_i \in \partial\psi_i^*(\hat{w}_i) \text{ for some } \hat{w}_i \in \mathbb{R} \tag{5.10}$$

by [12, Theorem 23.5]. Therefore W_i is nonempty. Now for any $w_i^1, w_i^2 \in W_i$, there exist $v_i^1, v_i^2 \in [v_i^-, v_i^+]$ such that $v_i^1 \in \partial\psi_i^*(w_i^1)$ and $v_i^2 \in \partial\psi_i^*(w_i^2)$, which implies $w_i^1, w_i^2 \in \mathrm{dom}(\partial\psi_i^*)$. For any $w_i^3 \in \,]w_i^1, w_i^2[$, it is obvious that $w_i^3 \in \mathrm{ri}\big(\mathrm{dom}(\partial\psi_i^*)\big)$. Hence $\partial\psi_i^*(w_i^3) \neq \emptyset$, and according to [12, Theorem 24.1], there holds $v_i^1 \leq v_i^3 \leq v_i^2$ for any $v_i^3 \in \partial\psi_i^*(w_i^3)$, which implies $w_i^3 \in W_i$.

In the following, we first prove the middle part of (5.8) for four different cases:

Case 1: both v_i^- and v_i^+ are finite. Then for any $w_i \in \,]w_i^-, w_i^+[$, we have $w_i \in W_i$. Hence $[v_i^-, v_i^+] \cap \partial\psi_i^*(w_i)$ is nonempty, i.e.,

$$[v_i^-, v_i^+] \cap \operatorname*{argmax}_{v_i \in \mathbb{R}} \{w_i v_i - \psi_i(v_i)\} \neq \emptyset.$$

Therefore

$$\begin{aligned} \psi_i^*(w_i) = \max_{v_i \in \mathbb{R}} \{w_i v_i - \psi_i(v_i)\} &= \max_{v_i^- \leq v_i \leq v_i^+} \{w_i v_i - \psi_i(v_i)\} \\ &= \rho_{v_i^-, v_i^+, \psi_i}(w_i) \quad \forall w_i \in \,]w_i^-, w_i^+[. \end{aligned} \tag{5.11}$$

Case 2: v_i^- is finite while $v_i^+ = +\infty$. Then $w_i^+ = +\infty$ by (5.7). For any $w_i \in \,]w_i^-, \hat{w}_i]$ with $\hat{w}_i$ defined in (5.10), there holds $w_i \in W_i$, and the argument to prove $\rho_{v_i^-, v_i^+, \psi_i}(w_i) = \psi_i^*(w_i)$ is similar to that in Case 1. Now consider $w_i > \hat{w}_i$. It follows from (5.10) that $0 \in \hat{w}_i - \partial\psi_i(\hat{v}_i)$. Hence the expression $\hat{w}_i v_i - \psi_i(v_i)$ as a concave function in v_i is nondecreasing on $]-\infty, \hat{v}_i]$. Then for any $w_i > \hat{w}_i$, the expression $w_i v_i - \psi_i(v_i)$ as a concave function in v_i is also nondecreasing on $]-\infty, \hat{v}_i]$. Therefore

$$\psi_i^*(w_i) = \sup_{v_i \in \mathbb{R}} \{w_i v_i - \psi_i(v_i)\} = \sup_{v_i^- \leq v_i} \{w_i v_i - \psi_i(v_i)\} = \rho_{v_i^-, v_i^+, \psi_i}(w_i) \quad \forall w_i > \hat{w}_i.$$

Case 3: $v_i^- = -\infty$, while v_i^+ is finite. The proof of this case is similar to that of Case 2.

Case 4: $v_i^- = -\infty$, and $v_i^+ = +\infty$. In this case, the conclusion follows directly from the definition of $\rho_{v_i^-, v_i^+, \psi_i}(w_i)$.

Next consider $w_i > w_i^+$ when $w_i^+ < +\infty$. Observe that in such a case, v_i^+ is finite by (5.7). We have shown in the argument of Case 2 above that for any $w_i > \hat{w}_i$, the expression $w_i v_i - \psi_i(v_i)$ as a concave function in v_i is nondecreasing on $]-\infty, \hat{v}_i]$. Hence the supremum of $w_i v_i - \psi_i(v_i)$ on $v_i \in [v_i^-, v_i^+]$ will be reached somewhere in $[\hat{v}_i, v_i^+]$. Note that $\hat{v}_i > v_i^-$. If the supremum is attained on any $\tilde{v}_i \in [\hat{v}_i, v_i^+[$, then $\tilde{v}_i \in \operatorname{argmax}_{v_i}\{w_i v_i - \psi_i(v_i)\}$, or $\tilde{v}_i \in \partial\psi_i^*(w_i)$, which is a contradiction to the definition of w_i^+. Therefore

$$\operatorname*{argmax}_{v_i^- \leq v_i \leq v_i^+} \{w_i v_i - \psi_i(v_i)\} = \{v_i^+\} \quad \forall w_i > w_i^+. \tag{5.12}$$

Similarly, we can prove

$$\operatorname*{argmax}_{v_i^- \leq v_i \leq v_i^+} \{w_i v_i - \psi_i(v_i)\} = \{v_i^-\} \quad \forall w_i < w_i^-. \tag{5.13}$$

It follows from (5.12) and (5.13) that

$$\rho_{v_i^-, v_i^+, \psi_i}(w_i) = w_i v_i^+ - \psi_i(v_i^+) \text{ for all } w_i > w_i^+, \tag{5.14}$$

$$\rho_{v_i^-, v_i^+, \psi_i}(w_i) = w_i v_i^- - \psi_i(v_i^-) \text{ for all } w_i < w_i^-. \tag{5.15}$$

It is obvious from (5.6) that $\rho_{v_i^-, v_i^+, \psi_i}$ is a convex function. Therefore $\rho_{v_i^-, v_i^+, \psi_i}$ is continuous in the relative interior of its effective domain. Now if $w_i^- < w_i^+$, then (5.8) follows from (5.14), (5.15) and (5.11) by elementary calculus. Otherwise if $w_i^- = w_i^+$, then $\hat{w}_i = w_i^- = w_i^+$. Recall that $\hat{v}_i \in]v_i^-, v_i^+[$ in (5.9). Hence

$$\psi_i^*(\hat{w}_i) = \hat{w}_i \hat{v}_i - \psi_i(\hat{v}_i) = \sup_{v_i^- \leq v_i \leq v_i^+} \{\hat{w}_i v_i - \psi_i(v_i)\} = \rho_{v_i^-, v_i^+, \psi_i}(\hat{w}_i),$$

and (5.8) still follows from (5.14) and (5.15) by elementary calculus. $\square$

With the aid of Proposition 5.1, various ρ terms can be derived by using conjugate pairs of convex functions. Hence the formulation in this paper not only allows for the merit function φ_i in the objective (5.5) being more general than linear-quadratic, but also provides a much wider variety for the monitoring functions ρ in modeling the constraints.

Example 1:

$$\psi_i(v_i) = \frac{1}{\alpha}|v_i|^\alpha \text{ and } \psi_i^*(w_i) = \frac{1}{\beta}|w_i|^\beta$$

are a conjugate pair [12], where $1 < \alpha < +\infty$, $1 < \beta < +\infty$ and $\frac{1}{\alpha} + \frac{1}{\beta} = 1$. Let $0 = v_i^- < v_i^+ < +\infty$. Then $w_i^- = 0$ and $w_i^+ = (v_i^+)^{\frac{1}{(\beta-1)}}$. Therefore (5.8) becomes

$$\rho_{0,v_i^+,\psi_i}(w_i) = \begin{cases} v_i^+(w_i - w_i^+) + \frac{1}{\beta}(w_i^+)^\beta & \text{if } w_i^+ \le w_i, \\ \frac{1}{\beta}(w_i)^\beta & \text{if } 0 < w_i < w_i^+, \\ 0 & \text{if } w_i \le 0. \end{cases}$$

Hence the ρ term will give any positive $w_i < w_i^+$ a penalty of the form $\frac{1}{\beta}(w_i)^\beta$ until w_i reaches w_i^+. After passing w_i^+, the penalty increases linearly with slope v_i^+. For $\beta = 2$, the ρ term reduces to the piecewise linear-quadratic monitoring function in ELQP. But the formulation here provides the whole variety of monitoring functions of w_i to the βth power with $\beta \in (1, +\infty)$.

Example 2:

$$\psi_i(v_i) = \begin{cases} v_i \log v_i - v_i & \text{if } v_i > 0, \\ 0 & \text{if } v_i = 0, \\ +\infty & \text{if } v_i < 0, \end{cases} \quad \text{and } \psi_i^*(w_i) = e^{w_i}$$

are a conjugate pair [12]. Let v_i^- and v_i^+ be such that $0 < v_i^- < v_i^+ \le +\infty$. Then $w_i^- = \log v_i^-$, $w_i^+ = \log v_i^+$ and (5.8) becomes

$$\rho_{v_i^-,v_i^+,\psi_i}(w_i) = \begin{cases} v_i^+(w_i - w_i^+) + v_i^+ & \text{if } w_i^+ \le w_i, \\ e^{w_i} & \text{if } w_i^- < w_i < w_i^+, \\ v_i^-(w_i - w_i^-) + v_i^- & \text{if } w_i \le w_i^-. \end{cases}$$

Hence the ρ term will give rise to an exponentially increasing penalty between $w_i^- = \log v_i^-$ and $w_i^+ = \log v_i^+$.

Proposition 5.2 (general properties of the ρ terms)**.** *The function $\rho_{V,\psi}$ in (5.3) is lower semicontinuous and convex with a nonempty effective domain. The same holds for $\rho_{U,\varphi}$.*

Proof. The function $\rho_{V,\psi}$ is in fact the conjugate of $\tilde{\psi} = \psi + \delta_V$, where δ_V is the indicator of V. Note that $\tilde{\psi}$ is a proper closed convex function by [12, Theorem 9.3] since $V \subset \operatorname{dom} \psi$ is nonempty. Hence the conclusion for $\rho_{V,\psi}$ follows from [12, Theorem 12.2]. The conclusion for $\rho_{U,\varphi}$ can be proved similarly. $\square$

Let "epi" denote the epigraph, and let "rc" denote the recession cone of either a convex set or a convex function (see [12]). The next proposition gives the condition for the monitoring functions defined in (5.3) to be finite-valued.

Proposition 5.3 (finiteness conditions for the ρ terms)**.** *Let $\pi_p : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ and $\pi_d : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^m$ be the canonical projections*

$$(u, \beta) \to u \quad \forall u \in \mathbb{R}^n,\ \beta \in \mathbb{R} \qquad \textit{and} \qquad (v, \gamma) \to v \quad \forall v \in \mathbb{R}^m,\ \gamma \in \mathbb{R}$$

respectively. If

$$\pi_d\big(\operatorname{rc}(\operatorname{epi} \psi)\big) \cap \operatorname{rc} V = \{0\}, \tag{5.16}$$

then $\operatorname{dom} \rho_{V,\psi} = \mathbb{R}^m$ *and*

$$\operatorname*{argmax}_{v \in V} \{r \cdot v - \psi(v)\} \neq \emptyset \quad \forall r \in \mathbb{R}^m. \tag{5.17}$$

Similarly, if

$$\pi_p\big(\operatorname{rc}(\operatorname{epi} \varphi)\big) \cap \operatorname{rc} U = \{0\}, \tag{5.18}$$

then $\operatorname{dom} \rho_{U,\varphi} = \mathbb{R}^n$ *and*

$$\operatorname*{argmax}_{u \in U} \{s \cdot u - \varphi(u)\} \neq \emptyset \quad \forall s \in \mathbb{R}^n. \tag{5.19}$$

Proof. For any $r \in \mathbb{R}^m$, define

$$\Psi_r(\,\cdot\,) = \psi(\,\cdot\,) - \langle r, \cdot \rangle. \tag{5.20}$$

We claim that (5.16) implies

$$\operatorname{rc} \Psi_r \cap \operatorname{rc} V = \{0\} \quad \forall r \in \mathbb{R}^m.$$

Then $\operatorname{dom} \rho_{V,\psi} = \mathbb{R}^m$, and (5.17) holds by [12, Theorem 27.3].

To prove the claim, we observe that $\operatorname{epi} \Psi_r$ is a nonempty closed convex set. Hence for any $(v, \gamma) \in \operatorname{rc}(\operatorname{epi} \Psi_r)$, there exist $\{\alpha^\mu, v^\mu, \gamma^\mu\}$ with $(v^\mu, \gamma^\mu) \in \operatorname{epi} \Psi_r$, $\alpha^\mu > 0$ and $\alpha^\mu \downarrow 0$, such that

$$(v, \gamma) = \lim_{\mu \to \infty} \alpha^\mu (v^\mu, \gamma^\mu) \tag{5.21}$$

by [12, Theorem 8.2]. Let

$$\tilde{\gamma}^\mu = \gamma^\mu + \langle r, v^\mu \rangle \text{ and } \tilde{\gamma} = \gamma + \langle r, v \rangle.$$

Then $(v^\mu, \tilde{\gamma}^\mu) \in \operatorname{epi} \psi$, and it follows from (5.21) that

$$(v, \tilde{\gamma}) = \lim_{\mu \to \infty} \alpha^\mu (v^\mu, \tilde{\gamma}^\mu).$$

Hence $(v, \tilde{\gamma}) \in \operatorname{rc}(\operatorname{epi} \psi)$ by [12, Theorem 8.2]. Therefore

$$\pi_d\big(\operatorname{rc}(\operatorname{epi} \Psi_r)\big) \subset \pi_d\big(\operatorname{rc}(\operatorname{epi} \psi)\big).$$

But $\operatorname{rc} \Psi_r \subset \pi_d\big(\operatorname{rc}(\operatorname{epi} \Psi_r)\big)$ according to the definition of the recession cone of convex functions [12]. Therefore the claim is true. The conclusion for $\rho_{U,\varphi}$ can be proved similarly. $\square$

The conditions in (5.16) and (5.18) are also sufficient for the existence of a saddle point for the Lagrangian (5.1) over $U \times V$.

Proposition 5.4 (existence of a saddle point)**.** *If both (5.16) and (5.18) in Proposition 5.3 hold, then the saddle value of $L(u, v)$ in (5.1) over $U \times V$ is finite, and there exists a saddle point $(\bar{u}, \bar{v})$ of $L(u, v)$ over $U \times V$.*

Proof. The function Ψ_r defined in (5.20) has the same effective domain as ψ. Hence by the inclusion in (5.2), it is obvious that

$$\operatorname{epi}(\Psi_r + \delta_V) = \operatorname{epi} \Psi_r \cap (V \times \mathbb{R}) \neq \emptyset.$$

Thus

$$\operatorname{rc}(\operatorname{epi}(\Psi_r + \delta_V)) = \operatorname{rc}(\operatorname{epi} \Psi_r) \cap \operatorname{rc}(V \times \mathbb{R})$$

by [12, Corollary 8.3.3]. Therefore

$$\pi_d\big(\mathrm{rc}(\mathrm{epi}(\Psi_r + \delta_V))\big) \subset \pi_d\big(\mathrm{rc}(\mathrm{epi}\,\Psi_r)\big) \cap \pi_d\big(\mathrm{rc}(V \times \mathbb{R})\big).$$

But $\pi_d\big(\mathrm{rc}(V \times \mathbb{R})\big) = \mathrm{rc}\,V$. Moreover we have already shown that

$$\pi_d\big(\mathrm{rc}(\mathrm{epi}\,\Psi_r)\big) \subset \pi_d\big(\mathrm{rc}(\mathrm{epi}\,\psi)\big) \quad \forall \Psi_r = \psi - \langle r, \,\cdot\,\rangle,\, r \in \mathbb{R}^m$$

in the proof of Proposition 5.3. Hence it follows from (5.16) that

$$\pi_d\big(\mathrm{rc}(\mathrm{epi}(\Psi_r + \delta_V))\big) = \{0\},$$

which implies that for any fixed u, the function $-L(u, \cdot) + \delta_V(\cdot)$ has no direction of recession. Similarly (5.18) implies that for any fixed v, the function $L(\cdot, v) + \delta_U(\cdot)$ has no direction of recession. Thus the conclusion of the proposition follows directly from Theorem 1.2. $\square$

Acknowledgments. The author thanks two anonymous referees for their very helpful comments and suggestions.

REFERENCES.

1. C. Zhu and R. T. Rockafellar, "Primal-dual projected gradient algorithms for extended linear-quadratic programming", SIAM J. Optimization 3 (1993), pp. 751–783.

2. C. Zhu, "On the primal-dual steepest descent algorithm for extended linear-quadratic programming", technical report, Department of Mathematical Sciences, The Johns Hopkins University, 1992. Accepted for publication by SIAM J. Optimization, 1993.

3. R. T. Rockafellar and R. J-B Wets, "A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming," Math. Programming Studies 28 (1986), pp. 63–93.

4. R. T. Rockafellar and R. J-B Wets, "Linear-quadratic problems with stochastic penalties: the finite generation algorithm," in *Numerical Techniques for Stochastic Optimization Problems*, Y. Ermoliev and R. J-B Wets (eds.), Springer-Verlag Lecture Notes in Control and Information Sciences No. 81 (1987), pp. 545–560.

5. R. T. Rockafellar, "A generalized approach to linear-quadratic programming," in *Proc. International Conf. on Numerical Optimization and Appl. (Xi'an, China, 1986)*, pp. 58–66.

6. R. T. Rockafellar, "Linear-quadratic programming and optimal control," SIAM J. Control Opt. 25 (1987), pp. 781–814.

7. R. T. Rockafellar and R. J-B Wets, "Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time," SIAM J. Control Opt. 28 (1990), pp. 810–822.

8. R. T. Rockafellar, "Computational schemes for solving large-scale problems in extended linear-quadratic programming," Math. Programming 48 (1990), pp. 447–474.

9. R. T. Rockafellar, "Large-scale extended linear-quadratic programming and multistage optimization," in *Proc. Fifth Mexico-U.S. Workshop on Numerical Analysis* (S. Gomez, J.-P. Hennart and R. Tapia, eds.), SIAM, 1990.

10. A. King, "An implementation of the Lagrangian finite generation method,"

in *Numerical Techniques for Stochastic Programming Problems*, Y. Ermoliev and R. J-B Wets (eds.), Springer-Verlag, 1988.

11. J. M. Wagner, *Stochastic Programming with Recourse Applied to Groundwater Quality Management*, doctoral dissertation, MIT, 1988.

12. R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, N.J., 1970.

13. G. H.-G. Chen and R. T. Rockafellar, "Forward-backward splitting methods in Lagrangian optimization," technical report, Department of Applied Mathematics University of Washington, May 1992.

14. C. Zhu, "Asymptotic convergence analysis of the splitting algorithm," technical report, Department of Mathematical Sciences, The Johns Hopkins University, 1991. Accepted for publication by Math. of Operations Research, 1993.

15. C. Zhu, "Modified proximal point algorithm for extended linear-quadratic programming," Computational Optimization and Application, 1 (1992), pp. 185–205.

16. J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables,* Academic Press, New York, 1970.