ELSEVIER

Operations Research Letters 19 (1996) 105–113

operations research letters

Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs

Hanif D. Sherali[a,*], Gyunghyun Choi[b]

[a] *Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0118, USA*
[b] *SI Consulting Team, Samsung Data Systems, Imkwang Bldg. 219-1, Migun-Dong, Seodaemun-Gu, Seoul, South Korea 120-020*

Received 1 November 1994; revised 1 January 1996

Abstract

Lagrangian duality is a frequently used technique for solving specially structured linear programs or for solving linear programming relaxations of nonconvex discrete or continuous problems within a branch-and-bound approach. In such cases, subgradient optimization methods provide a valuable tool for obtaining quick solutions to the Lagrangian dual problem. However, little is known or available for directly obtaining primal solutions via such a dual optimization process without resorting to penalty functions, or tangential approximation schemes, or the solution of auxiliary primal systems. This paper presents a class of procedures to recover primal solutions directly from the information generated in the process of using pure or deflected subgradient optimization methods to solve such Lagrangian dual formulations. Our class of procedure is shown to subsume two existing schemes of this type that have been proposed in the context of pure subgradient approaches under restricted step size strategies.

Keywords: Lagrangian duality; Lagrangian relaxation; Linear programming; Deflected subgradient optimization; Primal recovery schemes

1. Introduction

Consider a linear programming problem stated in the form;

$$\text{LP:}\quad \text{Minimize}\quad cx \qquad \text{subject to}\quad Ax \leqslant b, \quad x \in X, \tag{1}$$

where A is $m \times n$, and where X is a nonempty polytope in R^n. We assume that X is specially structured so that it is relatively easy to solve linear programming problems over simply this set X. Accordingly, we might choose to adopt a price-directive Lagrangian relaxation approach (see [5, 4]) in which we can solve LP via the Lagrangian dual problem

$$\text{LD:}\quad \text{Maximize}\ \{\theta(\pi) \colon \pi \geqslant 0\}, \tag{2a}$$

where $\theta(\pi)$ is evaluated via the *Lagrangian subproblem*

$$\text{LS}(\pi)\colon\ \theta(\pi) = \text{Minimum}\ \{cx + \pi^t(Ax - b) \colon x \in X\}. \tag{2b}$$

* Corresponding author.

PII S0167-6377(96)00019-3

It is well-known that $\theta(\cdot)$ defined by (2b) is concave and piecewise linear, and that LD can be solved via a cutting plane or tangential approximation approach to recover both primal and dual optimal solutions to LP (see [2] for example). However, a more popular approach to solve the nondifferentiable optimization Problem LD is by subgradient optimization techniques. This is particularly useful when such problems arise via linear programming relaxations LP of nonconvex discrete or continuous optimization problems, and several instances of these relaxations might need to be solved within the framework of a branch-and-bound approach. In such cases, subgradient optimization techniques can provide a significant computational advantage over simplex based or interior point LP solvers (for example, see [4] and the computational experience reported in Adams and Sherali, [1] and Sherali and Tuncbilek, [11, 12]. The reason for this is the quick and simple manner in which an optimum (or practically, a near-optimum) can be found for LD via this approach. One such scheme that employs pure subgradient based search directions might proceed as follows. Given any iterate $\pi_k \geqslant 0$, $k \geqslant 1$, a subgradient g_k of $\theta(\cdot)$ at $\pi = \pi_k$ is computed as

$$g_k = Ax_{\pi_k} - b, \text{ where } x_{\pi_k} \in X_{\pi_k} \equiv \{x\colon x \text{ solves for } \theta(\pi_k) \text{ via (2b)}\}. \tag{3}$$

Then, the new iterate π_{k+1} is determined by

$$(\pi_{k+1})_i = \text{maximum}\{0, (\pi_k + \lambda_k g_k)_i\} \quad \text{for each } i = 1, \dots, m, \tag{4}$$

where $\lambda_k > 0$ is a suitable step length.

While such a subgradient optimization approach can be quite powerful in providing a quick lower bound on LP via the solution of LD, the disadvantage is that a primal optimal solution (or even a feasible, near-optimal solution) to LP is not usually available via this scheme. Such a primal solution is of importance not only when LP itself is the primary problem of interest, but also in the context of branch-and-bound approaches where a solution to the linear programming relaxation LP might be required, as for example, in order to generate a feasible solution to the underlying nonconvex problem, or to implement partitioning decisions, or to generate valid inequalities. Moreover, the availability of a primal solution can provide a natural stopping criterion based on the duality gap for the subgradient optimization procedure that otherwise lacks such a feature. Sen and Sherali [9] and Sherali and Ulular [13] show how a primal penalty function can be coordinated with the Lagrangian dual problem in a primal-dual subgradient optimization approach. However, this now requires the optimization to be conducted in the joint primal-dual space. Alternatively, one can attempt to obtain a primal feasible solution after enforcing complementary slackness on the derived dual optimal solution. But again, this might involve a significant additional computational burden, particularly in the case of dual degeneracy. On the other hand, by adopting certain restricted step size strategies, and accumulating a specific convex combination of the primal *subproblem* solutions x_{π_k} generated via (3) for $k = 1, 2, \dots,$ Shor [14] and Larsson and Liu [7] show how one can recover a primal optimal solution while solving LD via a *pure subgradient approach* (given by (3) and (4)), using only the information generated in this process itself. This is the type of primal solution recovery scheme that is of interest to us in this paper.

In Section 2 below, we present a convex combination weighting strategy to recover primal feesible and optimal solutions that permits a more flexible choice of step lengths. This is important since the choice of step lengths can greatly influence the computational efficiency of subgradient methods. As we demonstrate, both the procedures of Shor [14] and Larsson and Liu [7] turn out to be special cases of our general result. This result is then extended in Section 3 to the case where deflected subgradient, rather than pure subgradient, directions are employed. This type of a strategy is known to be useful in alleviating the phenomenon of zig-zagging (see [3, 13]) and we show how primal solutions are recoverable when employing any such strategy as well. Several related open questions are posed in regard to theoretical convergence issues dealing with both dual as well as primal-dual approaches.

2. Recovery of primal solutions in pure subgradient approaches

Consider Problem LP and its Lagrangian dual LD given by (1) and (2), respectively, and suppose that LD is solved using a (pure) subgradient approach of the type (3),(4), under some suitable/desirable rule for selecting the step lengths $\lambda_k > 0$, $\forall k$. At each iteration k, following Shor [14], we will compose a primal iterate x_k via:

$$x_k = \sum_{j=1}^{k} \mu_j^k x_{\pi_j}, \text{ where } \sum_{j=1}^{k} \mu_j^k = 1 \text{ and } \mu_j^k \geqslant 0 \quad \text{for } j = 1, \ldots, k. \tag{5}$$

Here, x_{π_j} is defined in (3) $\forall j$, and μ_j^k $\forall j = 1, \ldots, k$ are some convex combination weights being used at iteration k. Hence, each x_k is a particular convex combination of the optimal solutions obtained for the Lagrangian dual subproblems $\mathrm{LS}(\pi_j)$ given by (2b), for $j = 1, \ldots, k$. In addition, let us define

$$\gamma_{jk} = \mu_j^k/\lambda_j, \quad j = 1, \ldots, k \quad \text{for each } k = 1, 2, \ldots . \tag{6}$$

and let

$$\Delta\gamma_k^{\max} \equiv \text{maximum}\{\gamma_{jk} - \gamma_{(j-1)k} \colon j = 2, \ldots, k\}. \tag{7}$$

Now, consider the following results.

Theorem 1. *Suppose that the subgradient method (3)–(4) operated with a suitable step length rule attains dual convergence to some feasible solution, that is, $\pi_k \to \bar{\pi}$ as $k \to \infty$ for some $\bar{\pi} \geqslant 0$. If the step lengths λ_k and the convex combination weights μ_j^k, $\forall j, k$, are chosen to satisfy:*

(i) $\gamma_{jk} \geqslant \gamma_{(j-1)k}$ *for all* $j = 2, \ldots, k$, *for each* k,

(ii) $\Delta\gamma_k^{\max} \to 0$ *as* $k \to \infty$, *and*

(iii) $\gamma_{1k} \to 0$ *as* $k \to \infty$ *and* $\gamma_{kk} \leqslant \delta$ *for all* k, *for some* $\delta > 0$,

then any accumulation point $\bar{x}$ of the sequence $\{x_k\}$ generated via (5) is feasible to Problem LP.

Proof. By Eqs. (3) and (5), we have

$$Ax_k - b = \sum_{j=1}^{k} \mu_j^k [Ax_{\pi_j} - b] = \sum_{j=1}^{k} \mu_j^k g_j,$$

where g_j is a subgradient of θ at π_j. Since $\pi_{j+1} \geqslant \pi_j + \lambda_j g_j$ by (4), we have from (8) that

$$\begin{aligned}
Ax_k - b &\leqslant \sum_{j=1}^{k} \mu_j^k \frac{1}{\lambda_j}(\pi_{j+1} - \pi_j) \\
&= -\frac{\mu_1^k}{\lambda_1}\pi_1 + \sum_{j=2}^{k}\left(\frac{\mu_{j-1}^k}{\lambda_{j-1}} - \frac{\mu_j^k}{\lambda_j}\right)\pi_j + \frac{\mu_k^k}{\lambda_k}\pi_{k+1} \\
&= -\gamma_{1k}\pi_1 - \sum_{j=2}^{k}(\gamma_{jk} - \gamma_{(j-1)k})\pi_j + \gamma_{kk}\pi_{k+1} \qquad (9) \\
&= -\gamma_{1k}\pi_1 + \gamma_{kk}\pi_{k+1} - \bar{\pi}\sum_{j=2}^{k}(\gamma_{jk} - \gamma_{(j-1)k}) + \sum_{j=2}^{k}(\gamma_{jk} - \gamma_{(j-1)k})(\bar{\pi} - \pi_j) \\
&= \gamma_{kk}(\pi_{k+1} - \bar{\pi}) + \gamma_{1k}(\bar{\pi} - \pi_1) + \sum_{j=2}^{k}(\gamma_{jk} - \gamma_{(j-1)k})(\bar{\pi} - \pi_j). \qquad (10)
\end{aligned}$$

Let v_k denote the last term in (10), and let us show that $v_k \to 0$ as $k \to \infty$. Toward this end, given any $\varepsilon > 0$, let $J \geqslant 2$ be large enough so that $\|\bar{\pi} - \pi_j\| \leqslant \varepsilon/2\delta$ for all $j > J$. Then, for $k \geqslant J$ and large enough so that, noting conditions (ii) and (iii) of the theorem, $\Delta\gamma_k^{\max}\sum_{j=2}^{J}\|\bar{\pi} - \pi_j\| \leqslant \varepsilon/2$ and $\gamma_{kk} \leqslant \delta$, we get by (i) and the triangle inequality that

$$\begin{aligned}
\|v_k\| &\leqslant \sum_{j=2}^{k}(\gamma_{jk} - \gamma_{(j-1)k})\|\bar{\pi} - \pi_j\| \\
&\leqslant \Delta\gamma_k^{\max}\sum_{j=2}^{J}\|\bar{\pi} - \pi_j\| + \frac{\varepsilon}{2\delta}\sum_{j=J+1}^{k}(\gamma_{jk} - \gamma_{(j-1)k}) \\
&\leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2\delta}(\gamma_{kk} - \gamma_{Jk}) \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Hence, since $\varepsilon > 0$ was arbitrary, we deduce that $v_k \to 0$ as $k \to \infty$. Moreover, since $\{x_k\}$ is contained in the compact set X, there exists a convergent subsequence $\{x_k\}_K \to \bar{x}$ indexed by some set K. Taking limits in (10) as $k \to \infty$, $k \in K$, we get upon using (iii) and that the last term approaches zero that $A\bar{x} \leqslant b$. Furthermore, since by (3) and (5), $x_k \in X$ for all k, we conclude that $\bar{x} \in X$ as well, and this completes the proof. □

Corollary 1. *Suppose that the sequence $\{\pi_k\}$ generated by the subgradient method is simply bounded, and that condition* (i) *of Theorem* 1 *holds along with the condition that $\gamma_{kk} \to 0$ as $k \to \infty$. Then any accumulation point $\bar{x}$ of the sequence $\{x_k\}$ generated via* (5) *is feasible to Problem LP.*

Proof. Eq. (9) and condition (i) implies that $Ax_k - b \leqslant \gamma_{kk}\pi_{k+1}\ \forall k$. Hence, by the hypothesis of the theorem, given that $\{x_k\}_K \to \bar{x}$, we have that $A\bar{x} - b \leqslant 0$, and since $\bar{x} \in X$ as well, the proof is complete. □

Theorem 1 above asserts that so long as the dual iterates are convergent, and the step lengths and the convex combination weights satisfy the stated conditions (i)–(iii), the corresponding sequence of primal iterates will produce a feasible solution, given that we keep track of the solution having a least measure of infeasibility. Corollary 1 asserts that even if the dual iterates are nonconvergent but remain bounded, the same result holds provided we select the parameters as stated. Note that in this case, condition (ii) and the condition $\{\gamma_{1k}\} \to 0$ in (iii) are implied by restricting the second condition in (iii) to require that $\{\gamma_{kk}\} \to 0$.

We now address the issue of primal and dual *optimality* within this scheme and show that under the conditions of Theorem 1, we actually obtain a pair of primal and dual optimal solutions.

Theorem 2. *Suppose that the subgradient method* (3)–(4) *operated with a suitable step length rule attains dual convergence $\{\pi_k\} \to \bar{\pi}$ to some feasible solution $\bar{\pi} \geqslant 0$, and let $\bar{x}$ be any accumulation point of the corresponding sequence of primal solutions $\{x_k\}$ generated via* (5). *If the step lengths λ_k and the convex combination weights μ_j^k, $\forall j, k$ satisfy the conditions* (i)–(iii) *of Theorem* 1, *then $\bar{x}$ and $\bar{\pi}$ are optimal solutions to the primal and dual problems LP and LD, respectively.*

Proof. Since θ is piecewise linear and concave, there exists a neighborhood $N(\bar{\pi})$ of $\bar{\pi}$ such that $\partial\theta(\pi) \subseteq \partial\theta(\bar{\pi})$ for all $\pi \in N(\bar{\pi})$, where $\partial\theta(\pi)$ is the subdifferential of θ at π. Hence, since $\{\pi_k\} \to \bar{\pi}$, we have from (3) that there exists a J_1 large enough so that $(Ax_{\pi_k} - b) \in \partial\theta(\pi_k) \subseteq \partial\theta(\bar{\pi})$ for all $k \geqslant J_1$. Moreover, from (3) and (4), there exists J_2 large enough so that for $k \geqslant J_2$,

$$(\pi_k)_i > 0 \text{ and } (Ax_{\pi_k} - b)_i = \frac{(\pi_{k+1})_i - (\pi_k)_i}{\lambda_k} \quad \text{if } (\bar{\pi})_i > 0. \tag{11}$$

Let us define $J = \text{maximum}\{J_1, J_2\}$, and consider the process once $k \geqslant J$.

Note that for $k \geqslant J$, using conditions (ii) and (iii) of the theorem, we have

$$\sum_{j=1}^{J} \gamma_{jk} \leqslant \gamma_{1k} + (J-1)\Delta\gamma_k^{\max} \to 0 \text{ as } k \to \infty.$$

From (6), this in turn means that $\mu_j^k \to 0$ $\forall j = 1, \ldots, J$ as $k \to \infty$, and so without loss of generality, we can assume that the procedure is re-initialized once $k \geqslant J$ gets large enough. Consequently, since we then have that each $x_{\pi_j} \in X_{\bar{\pi}}$ as defined in (3), we have from (5) that $x_k \in X_{\bar{\pi}}\ \forall k$, and so, $\bar{x} \in X_{\bar{\pi}}$ as well. Moreover, by Theorem 1, $\bar{x}$ is feasible to the primal problem. Hence, by the strong duality theorem (see [2, Theorem 6.2.5]), it remains to show that the complementary slackness condition holds true in order to complete the proof.

Toward this end, note that if $(\bar{\pi})_i > 0$ for any component i, we have from (5),(6), and (11) that for all k,

$$\begin{aligned}(Ax_k - b)_i &= \sum_{j=1}^{k} \mu_j^k (Ax_{\pi_j} - b)_i \\ &= \sum_{j=1}^{k} \frac{\mu_j^k}{\lambda_j}(\pi_{j+1} - \pi_j)_i \\ &= \sum_{j=1}^{k} \gamma_{jk}(\pi_{j+1} - \pi_j)_i \\ &= -\gamma_{1k}(\pi_1)_i - \sum_{j=2}^{k} (\gamma_{jk} - \gamma_{(j-1)k})(\pi_j)_i \\ &\quad + \gamma_{kk}(\pi_{k+1})_i \\ &= \gamma_{kk}(\pi_{k+1} - \bar{\pi})_i + \gamma_{1k}(\bar{\pi} - \pi_1)_i \\ &\quad + \sum_{j=2}^{k} (\gamma_{jk} - \gamma_{(j-1)k})(\bar{\pi} - \pi_j)_i. \end{aligned} \tag{12}$$

Noting Eqs. (10) and (12) and the related argument used in the proof of Theorem 1, we again have that the expression in (12) converges to 0 as $k \to \infty$. Therefore, by taking limits as $k \to \infty$, $k \in K$ (where $\{x_k\}_K \to \bar{x}$), we have that $(A\bar{x} - b)_i = 0$ whenever $(\bar{\pi})_i > 0$, and hence the complementary slackness condition holds true. This completes the proof. □

Corollary 2 (Shor's [14] Rule). *Let the step lengths λ_k and the convex combination weights μ_j^k, $\forall j,k$, satisfy*

$$\lambda_k > 0, \quad \lim_{k\to\infty} \lambda_k = 0, \quad \sum_{k=1}^{\infty} \lambda_k = \infty \quad \text{and}$$

$$\mu_j^k = \lambda_j \Big/ \sum_{j=1}^{k} \lambda_j \quad \forall j = 1, \dots, k, \ \forall k. \tag{13}$$

Then conditions (i)–(iii) *of the theorem hold true.*

Proof. Note by (6) and (13) that $\gamma_{jk} = 1/\sum_{t=1}^{k} \lambda_t$ for all $j = 1, \dots, k$, for all k. Thus, we have $\gamma_{jk} = \gamma_{(j-1)k}$ for all $j = 2, \dots, k$, which in turn implies that $\Delta\gamma_k^{\max} \equiv 0$ for all k. Also, $\sum_{k=1}^{\infty} \lambda_k = \infty$ implies that $\gamma_{kk} \to 0$ as $k \to \infty$, and this completes the proof. □

Corollary 3 (Larsson and Liu's [7] Rule). *Let the step lengths λ_k and the convex combination weights μ_j^k, $\forall j,k$, be given by*

$$\lambda_k = a/(b + ck) \ \forall k, \quad \mu_j^k = 1/k \ \forall j = 1, \dots, k, \ \forall k, \tag{14}$$

where $a > 0$, $b \geq 0$, and $c > 0$ are some chosen numbers. Then conditions (i)–(iii) *of the theorem hold true.*

Proof. By (6) and (14), we have that $\gamma_{jk} = (b + cj)/ak$ for $j = 1, \dots, k$, for all k. Hence, $\gamma_{jk} - \gamma_{(j-1)k} = c/ak > 0$ for $j = 2, \dots, k$, and $\Delta\gamma_k^{\max} = c/ak \to 0$ as $k \to \infty$. Moreover, $\gamma_{1k} \to 0$ and $\gamma_{kk} \to c/a$ as $k \to \infty$. This completes the proof. □

Remark 1. Note that Larsson and Liu's *step length rule* given by (14) is a special case of Shor's step length rule given by (13). Clearly, $\lambda_k > 0$ $\forall k$ and $\lim_{k\to\infty} \lambda_k = 0$. Moreover, since $a/(b + ck) \geq a/(b + c)k$ $\forall k$, we have that $\sum_{k=1}^{\infty} \lambda_k \geq [a/(b + c)] \sum_{k=1}^{\infty} (1/k) = \infty$. However, the average *convex combination rule* in (14) appears to be more reasonable than that suggested by Shor because for each k, the rule in (13) assigns larger weights to earlier subproblem solutions generated since $\{\lambda_j\} \to 0$.

As an illustration, we now present an alternative choice of λ_k and μ_j^k, $\forall j,k$, that satisfy the conditions of Theorems 1 and 2.

Corollary 4. *Suppose that a step length strategy that satisfies*

$$0 < \lambda_{k+1} \leq \lambda_k \text{ for all } k, \text{ and } \lim_{k\to\infty} k\lambda_k = \infty \tag{15}$$

attains dual convergence, that is $\{\pi_k\} \to \bar{\pi}$ for some $\bar{\pi} \geq 0$. If the convex combination weights are given by the average weighting rule, $\mu_j^k = 1/k$, for $j = 1, \dots, k$, $\forall k$, then the assertions of Theorems 1 and 2 hold true. In particular, if the step lengths are given by

$$\lambda_k = k^{-\mu} \text{ for all } k, \text{ where } 0 < \mu < 1, \tag{16}$$

then dual convergence is attained and the condition (15) is satisfied.

Proof. Note by (6) that $\gamma_{jk} = 1/(k\lambda_j)$ for $j = 1, \dots, k$, $\forall k$. Thus, condition (i) holds true since $\lambda_j \leq \lambda_{j-1}$. Moreover, we have,

$$\gamma_{jk} - \gamma_{(j-1)k} = \frac{1}{k}\left(\frac{1}{\lambda_j} - \frac{1}{\lambda_{j-1}}\right) = \frac{1}{k}\frac{(\lambda_{j-1} - \lambda_j)}{\lambda_j\lambda_{j-1}} \leq \frac{1}{k}\frac{\lambda_{j-1}}{\lambda_j\lambda_{j-1}} = \frac{1}{k\lambda_j},$$

which implies that

$$\Delta\gamma_k^{\max} \leq \max_{2\leq j\leq k} \frac{1}{k\lambda_j} = \frac{1}{k\lambda_k} = \gamma_{kk}.$$

But by (15), we have that $\gamma_{kk} = 1/(k\lambda_k) \to 0$ as $k \to \infty$, and so, $\Delta\gamma_k^{\max} \to 0$ as $k \to \infty$. Furthermore, $\gamma_{1k} = (1/k\lambda_1) \to 0$ as $k \to \infty$. This proves the first part of the corollary. Now, if the step lengths are given by (16), then as in Polyak [8], convergence to a dual optimal solution is obtained since the step lengths satisfy (13). Moreover, (15) also holds true, and this completes the proof. □

Note that Shor [14] and Larsson and Liu [7] each prescribe a restricted step length rule that, in particular, guarantees convergence to a dual optimum, along with a specific convex combination weighting scheme in order to recover a primal optimal solution. As pointed out in Remark 1, Larsson and Liu's step length rule is a special case of that prescribed by Shor, and furthermore, Shor's proof can also be modified (by revising the derivation on p. 118, line 4) to accommodate the average convex combination weighting rule in (14), using either the step length strategy in (13) or that suggested in (15) of Corollary 4. In fact, conditions (i)–(iii) composed in Theorem 1 attempt to capture the essence of what is required to establish the fundamental saddle-point optimality criteria. These conditions permit a greater degree of flexibility in selecting step lengths, appropriately coordinated with a suitable convex combination weighting scheme, including heuristic step length rules that might not necessarily guarantee convergence to a dual optimum, but as Theorem 2 asserts, are sufficient to recover primal and dual optimal solutions given that dual convergence does occur. Moreover, in the case of dual nonconvergence, Corollary 1 provides conditions under which at least a primal feasible solution might be recovered.

Remark 2. Note that in practice, subgradient optimization methods are often used in conjunction with periodic restarts, wherein the procedure is reset to the current incumbent solution, and termination is frequently based on performing a fixed limit of iterations. In light of the foregoing discussion, it might be advisable to switch on the primal recovery scheme embodied by (5) and satisfying the conditions of Theorems 1 and 2 during the final 50–100 iterations, following such a restart. Moreover, if the selected primal recovery convex combination weighting scheme satisfies the condition that

$$\mu_j^{k+1}/\mu_j^k \equiv \phi_k \text{ is independent of } j, \text{ for each } k, \quad (17)$$

then the set of primal iterates generated over previous iterations during this primal recovery process need not be stored. Instead, we can adopt the following update procedure, using (5) and (17).

$$\begin{aligned} x_{k+1} &\equiv \sum_{j=1}^{k+1} \mu_j^{k+1} x_{\pi_j} \\ &= \sum_{j=1}^{k} \mu_j^{k+1} x_{\pi_j} + \mu_{k+1}^{k+1} x_{\pi_{k+1}} \\ &= \phi_k x_k + \mu_{k+1}^{k+1} x_{\pi_{k+1}}. \end{aligned}$$

In particular, note that (17) holds true for each of the schemes presented in Corollaries 2–4, where $\phi_k = (\sum_{j=1}^{k} \lambda_t / \sum_{t=1}^{k+1} \lambda_t)$ for Corollary 2, and $\phi_k = k/(k+1)$ for Corollaries 3 and 4. □

3. Recovery of primal solutions in deflected subgradient approaches

In this section, we will extend our results and analysis of the previous section to the case of deflected subgradient methods. For the sake of convenience, we will consider linear programs in the equality form (after adding slack variables if necessary),

$$\text{LP: Minimize } \{cx\colon Ax = b, \quad x \in X\}, \quad (18)$$

for which the Lagrangian dual is given as in (2), except that the Lagrange multiplier vector π is now unrestricted in sign. Our motivaton for considering this class of procedures is that it is well known that the pure subgradient direction can lead to a zigzagging phenomenon, inducing a slow tail-end convergence, while subgradient deflection strategies tend to alleviate this undesirable behavior (see [3, 13]). In such methods, the direction d_k at any iteration $k \geqslant 2$ is computed by deflecting the pure subgradient direction via the previous direction of motion d_{k-1} according to

$$d_k = g_k + \psi_k d_{k-1}, \quad \text{where } d_1 = g_1, \quad (19\text{a})$$

and where, as given by (3), g_k is a subgradient of $\theta(\cdot)$ at the current iterate π_k, and $\psi_k \geqslant 0$ is a suitable deflection parameter. The revised iterate π_{k+1} is then given by

$$\pi_{k+1} = \pi_k + \lambda_k d_k, \quad (19\text{b})$$

where $\lambda_k > 0$ is a suitable step length. As before, we will consider the recovery of primal solutions via the scheme embodied by (3) and (5). Consider the following results.

Lemma 1. *Suppose that the deflection parameter ψ_k, and the convex combination weights μ_j^k, $\forall j,k$, have been selected to satisfy*

$$\psi_j \mu_j^k = \mu_{j-1}^k \quad \text{for } j = 2, \dots, k, \ \forall k. \tag{20}$$

Then, with x_k defined by (3) *and* (5), *we have,*

$$Ax_k - b = \frac{\mu_k^k}{\lambda_k}(\pi_{k+1} - \pi_k) \quad \forall k. \tag{21}$$

Proof. For each k, since $g_k = d_k - \psi_k d_{k-1}$ and $d_k = (\pi_{k+1} - \pi_k)/\lambda_k$ by (19), we have using (3) and (5) that

$$\begin{aligned} Ax_k - b &= \sum_{j=1}^{k} \mu_j^k g_j \\ &= \frac{\mu_1^k}{\lambda_1}(\pi_2 - \pi_1) \\ &\quad + \sum_{j=2}^{k} \mu_j^k \left[\frac{1}{\lambda_j}(\pi_{j+1} - \pi_j) \right. \\ &\qquad \left. - \psi_j \frac{1}{\lambda_{j-1}}(\pi_j - \pi_{j-1}) \right] \\ &= \left(\frac{\mu_2^k}{\lambda_1}\psi_2 - \frac{\mu_1^k}{\lambda_1} \right) \pi_1 \\ &\quad + \sum_{j=2}^{k-1} \left(\frac{\mu_{j-1}^k}{\lambda_{j-1}} - \frac{\mu_j^k}{\lambda_j} \right. \\ &\qquad \left. - \frac{\mu_j^k}{\lambda_{j-1}}\psi_j + \frac{\mu_{j+1}^k}{\lambda_j}\psi_{j+1} \right) \pi_j \\ &\quad + \left(\frac{\mu_{k-1}^k}{\lambda_{k-1}} - \frac{\mu_k^k}{\lambda_k} - \frac{\mu_k^k}{\lambda_{k-1}}\psi_k \right) \pi_k + \frac{\mu_k^k}{\lambda_k}\pi_{k+1}. \end{aligned} \tag{22}$$

From (20), we notice that the coefficients of $\pi_1, \pi_2, \dots, \pi_{k-1}$ in (22) are zeros, and that the coefficient of π_k in (22) is simply $-\mu_k^k/\lambda_k$. Hence, (22) reduces to (21) and this completes the proof. □

Remark 3. Given any subgradient deflection scheme defined by some choice of the parameters ψ_k, $k \geq 2$, we can determine the convex combination weights μ_j^k, $\forall j = 1, \dots, k$, for each k, as follows. Applying (20) recursively and using the fact that $\sum_{j=1}^{k} \mu_j^k = 1$, we derive,

$$\mu_j^k = \left[\prod_{t=j+1}^{k} \psi_t \right] \mu_k^k \quad \forall j = 1, \dots, k-1,$$

$$\text{where } \mu_k^k = 1 \Bigg/ \left[1 + \sum_{j=1}^{k-1} \prod_{t=j+1}^{k} \psi_t \right]. \tag{23}$$

Then given that the deflection scheme guarantees dual convergence to some solution $\bar{\pi}$ and that the sequence $\{\mu_k^k/\lambda_k\}$ remains bounded, (21) of Lemma 1 asserts that any accumulation point $\bar{x}$ of the sequence $\{x_k\}$ thus generated would be feasible to LP. The following theorem addresses the issue of primal and dual optimality. □

Theorem 3. *Suppose that a deflected subgradient algorithm of the type* (19) *operated with a suitable step length rule attains dual convergence $\{\pi_k\} \to \bar{\pi}$ to some solution $\bar{\pi}$, and let $\bar{x}$ be any accumulation point of the corresponding sequence of primal solutions $\{x_k\}$ generated via* (5). *If the deflection parameters ψ_k, the step lengths λ_k and the convex combination weights μ_j^k, $\forall j,k$, satisfy* (20) *along with* (i) *$\gamma_{kk} \equiv \mu_k^k/\lambda_k \leq \delta$ for all k, for some $\delta > 0$, and* (ii) *$\mu_j^k \to 0$ $\forall j = 1,, \dots, J$, for any fixed J, as $k \to \infty$, then $\bar{x}$ and $\bar{\pi}$ are optimal solutions to the primal and dual problems LP and LD, respectively.*

Proof. By Lemma 1 and the stated condition (i) on γ_{kk}, we have from (21) by taking limits as $k \to \infty$, $k \in K$ (where $\{x_k\}_K \to \bar{x}$) that $A\bar{x} = b$. Moreover, by (3) and (5), since $x_k \in X$ $\forall k$, we have that $\bar{x} \in X$ as well. Hence, $\bar{x}$ is feasible to LP. Moreover, by condition (ii) and the proof of Theorem 2, we have that $\bar{x} \in X_{\bar{\pi}}$. Hence, by Theorem 6.2.5 in Bazaraa et al. [2], we conclude that $\bar{x}$ and $\bar{\pi}$ are a pair of primal and dual optimal solutions, and this completes the proof. □

Remark 4. Note that the parameter choices given by Corollaries 2–4 satisfy conditions (i) and (ii) of Theorem 3. Additionally, if the deflection parameters ψ_k are chosen to satisfy (20), then this would fulfill all the conditions required by Theorem 3. For the choice of convex combination weights in Corollary 2, we have from (20) that $\psi_k = \lambda_{k-1}/\lambda_k$ $\forall k \geq 2$.

This would mean from (19) that the correction $\lambda_k d_k$ applied to π_k in order to obtain π_{k+1} would be given by $\lambda_k d_k = \lambda_k g_k + \lambda_{k-1} d_{k-1}$, $\forall k \geqslant 2$. (This would be implementable if the prescribed step length formula for λ_k is independent of d_k, as in (14) or (15), for example.) In the case of Corollaries 3 and 4, $\psi_k \equiv 1$ $\forall k \geqslant 2$. This yields the search direction $d_k = g_k + d_{k-1}$, $\forall k \geqslant 2$. Note that this is related to the *average direction strategy* of Sherali and Ulular [13] in which each of g_k and d_{k-1} are also, in essence, normalized to be of unit lengths. Hence, *if* these foregoing choices of parameters (ψ, λ, μ) yield dual convergence, as they do for the pure subgradient approach, we would then be able to recover a pair of primal and dual optimal solutions. However, whether or not such a scheme would yield dual convergence is an open question.

Next, let us address known convergence results for subgradient deflection schemes applied in the context of solving LD given by (2), and how these relate to our results. Camerini et al. [3] show that given a target value $\bar{\theta} < \theta^*$, where θ^* is the optimal value for (2), if step lengths λ_k are selected to satisfy $\varepsilon[\bar{\theta} - \theta(\pi_k)]/\|d_k\|^2 \leqslant \lambda_k \leqslant [\bar{\theta} - \theta(\pi_k)]/\|d_k\|^2$, for some $0 < \varepsilon < 1$, and the deflection parameter is selected as $\psi_k = \gamma_k |d_{k-1}^t g_k| / \|d_{k-1}\|^2$ if $d_{k-1}^t g_k < 0$ and 0 otherwise, where $0 \leqslant \gamma_k \leqslant 2$ $\forall k$, then either finitely we would find a π_l such that $\theta(\pi_l) \geqslant \bar{\theta}$, or else, $[\pi_k] \to \bar{\pi}$, where $\theta(\bar{\pi}) = \bar{\theta}$. Based on the convergence arguments for the primal-dual algorithm presented in Sherali and Ulular [13], Sherali and Choi [10] have developed two other generalized convergence results that permit a greater flexibility in selecting the deflection parameter ψ, requiring it to be simply nonnegative. The first result states that if

$$\bar{\theta} \leqslant \theta^*,\ \lambda_k = \beta_k \frac{[\bar{\theta} - \theta(\pi_k)]}{\|d_k\|^2},\ \text{where } 0 < \bar{\varepsilon} \leqslant \beta_k \leqslant 1$$

$$\text{for some } \bar{\varepsilon} > 0, \text{ and if } \psi_k \geqslant 0 \quad \forall k, \tag{25}$$

then either $\theta(\pi_l) \geqslant \bar{\theta}$ for some iteration l, or else, $\lim_{k\to\infty} \theta(\pi_k) = \bar{\theta}$. The second result addresses the situation when no suitable target value $\bar{\theta}$ that satisfies $\bar{\theta} \leqslant \theta^*$, and is yet greater than any known incumbent value, might be practically available. In this case, if we select

$$\lambda_k = \frac{\varepsilon}{\|d_k\|^2} \text{ for any given } \varepsilon > 0,$$

$$\text{and if } \psi_k \geqslant 0 \quad \forall k, \tag{26}$$

then for some iteration l, we will attain $\theta(\pi_l) \geqslant \theta^* - \varepsilon$.

As far as the recovery of primal solutions within the context of these convergence results is concerned, note that the Camerini et al. [3] scheme precludes dual convergence to an exact optimum solution, and moreover, the application of condition (2) can be problematic as ψ_j can possibly be zero. (In fact, as pointed out by Sherali and Choi [10], ψ_j frequently turns out to be zero in computational runs, and the method often performs close to a pure subgradient algorithm.) On the other hand, conditions (25) and (26) provide a greater degree of flexibility in the choice of the deflection parameter, hence permitting the conditions of Theorem 3 to possibly hold. For example, suppose that we select $\psi_k = 1$ $\forall k$. Then from (20) and (23), we would have $\mu_j^k = 1/k$ $\forall j = 1, \ldots, k$, for each $k = 1, 2, \ldots$, so that condition (ii) of Theorem 3 holds true. Moreover, suppose that $\bar{\theta} = \theta^*$ in (25) (where the optimum might be known if we are dealing with primal-dual systems), and that $\{\pi_k\} \to \bar{\pi}$ in the dual convergence process, so that $\theta(\bar{\pi}) = \theta^*$. Then condition (i) of Theorem 3 would also be satisfied provided that $\|d_k\|^2/[\beta_k k(\theta^* - \theta(\pi_k))]$ remains bounded. This might happen if, for example, $\|d_k\|$ remains bounded from above, and if the indeterminate quantity $k[\theta^* - \theta(\pi_k)]$ remains bounded away from zero as $k \to \infty$. Admittedly, this is quite a stringent requirement.

On the other hand, in regard to condition (26), which only ensures ε-optimality, if we actually do realize dual convergence to optimality, and if $\|d_k\|^2/\varepsilon k$ remains bounded, then we would again have the conditions of Theorem 3 holding true. In this context of dual convergence under (26), there is another choice of deflection parameters ψ_k that would satisfy the conditions of Theorem 3. Consider $\psi_k = \psi$ $\forall k$, where $0 < \psi < 1$. Then, from (20) and (23), we have

$$\mu_k^k = \frac{(1-\psi)}{(1-\psi^k)} \quad \text{and} \quad \mu_j^k = \frac{(1-\psi)\psi^{k-j}}{(1-\psi^k)}$$
$$\forall j = 1, \dots, k, \quad \forall k. \tag{27}$$

Hence, noting from (19a) that $d_k = g_k + \psi g_{k-1} + \psi^2 g_{k-2} + \cdots + \psi^{k-1} g_1$, and that the norm of each subgradient is bounded by some quantity M as evident from Eq. (3), we have for each k that $\|d_k\| \leqslant M(1-\psi^k)/(1-\psi) \leqslant M/(1-\psi)$. Hence, we obtain from (26) and (27) that

$$\frac{\mu_k^k}{\lambda_k} = \frac{(1-\psi)\|d_k\|^2}{(1-\psi^k)\varepsilon} \leqslant \frac{\|d_k\|^2}{\varepsilon} \leqslant \frac{M^2}{(1-\psi)^2\varepsilon}. \tag{28}$$

From (27) and (28), we see that both conditions (i) and (ii) of Theorem 3 hold true. Hence, under (26), if we were to realize dual convergence to optimality using $\psi_k = \psi \ \forall k$, where $0 < \psi < 1$, then we would obtain a pair of primal and dual optimal solutions via the prescribed algorithmic process.

Finally, we also mention here that there exist convergence results for generalized subgradient procedures, which are related to deflected subgradient methods, under the divergent step size rule of the type (13) (see [6] for example). However, these results essentially require the procedure to asymptotically behave as a pure subgradient method. Hence, it appears possible to compose the results of Sections 2 and 3 to derive similar primal-dual convergence theorems for such generalized subgradient procedures.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. DMI9121419.

References

[1] W.P. Adams and H.D. Sherali, "Mixed-integer bilinear programming problems", *Math. Programming* **59**, 279–305, 1993.

[2] M.S. Bazaraa, H.D. Sherali and C.M. Shetty, *Nonlinear Programming: Theory and Algorithms.* 2nd ed., Wiley, New York, 1993.

[3] P.M. Camerini, L. Fratta and F. Maffioli, "On improving relaxation methods by modified gradient techniques", *Math. Programming Study* **3**, 26–34 (1975).

[4] M.L. Fisher, "The Lagrangian relaxation method for solving integer programming problems", *Management Sci.* **27**, 1–18 (1981).

[5] A.M. Geoffrion, "Lagrangian relaxation for integer programming", in: M.L. Balinski (ed.), *Mathematical Programming Study*, Vol. 2, North-Holland, Amsterdam, 1974, pp. 82–114.

[6] S. Kim and H. Ahn, "Convergence of a generalized subgradient method for nondifferentiable convex optimization", *Math. Programming* **50**, 75–80 (1991).

[7] T. Larsson and Z. Liu "A primal convergence result for dual subgradient optimization with application to multicommodity network flows", Research Report, Dept. of Math., Linkoping Institute of Technology, S-581 83 Linkoping, Sweden, 1989.

[8] B.T. Polyak 'A general method for solving extremum problems", *Soviet Math.* **8**, 593–597 (1967).

[9] S. Sen and H.D. Sherali, "A class of convergent primal-dual subgradient algorithms for decomposable convex programs", *Math. Programming* **35**, 279–297 (1986).

[10] H.D. Sherali and G. Choi, "A variable target value method for nondifferentiable optimization", Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0118, 1994.

[11] H.D. Sherali and C.H. Tuncbilek, "A squared-Euclidean distance location-allocation problem", *Naval Res. Logist.* **39**, 447–469 (1992).

[12] H.D. Sherali and C.H. Tuncbilek, " A reformulation–convexification approach for solving nonconvex quadratic programming problems", *J. Global Optim.* **7**, 1–31 (1995).

[13] H.D. Sherali and O. Ulular, "A primal-dual conjugate subgradient algorithm for specially structured linear and convex programming problems", *Appl. Math. Optim.* **20**, 193–221 (1989).

[14] N.Z. Shor, *Minimization Methods for Nondifferentiable Functions*, Translated from Russian by K.C. Kiwiel and A. Ruszczynski, Springer, Berlin, 1985, pp. 116–120.