INFORMS Journal on Computing

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# An Effective Deflected Subgradient Optimization Scheme for Implementing Column Generation for Large-Scale Airline Crew Scheduling Problems

Shivaram Subramanian, Hanif D. Sherali,

*INFORMS Journal on Computing*
Vol. 20, No. 4, Fall 2008, pp. 565–578
ISSN 1091-9856 | EISSN 1526-5528 | 08 | 2004 | 0565

DOI 10.1287/ijoc.1080.0267


# An Effective Deflected Subgradient Optimization Scheme for Implementing Column Generation for Large-Scale Airline Crew Scheduling Problems

Shivaram Subramanian
Enterprise Optimization, United Airlines (WHQKB), Elk Grove Village, Illinois 60007, shiva.subramanian@oracle.com

Hanif D. Sherali
Grado Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, hanifs@vt.edu

We present a new deflected subgradient scheme for generating good quality dual solutions for linear programming (LP) problems and utilize this within the context of large-scale airline crew planning problems that arise in practice. The motivation for the development of this method came from the failure of a black-box-type approach implemented at United Airlines for solving such problems using column generation in concert with a commercial LP solver, where the software was observed to stall while yet remote from optimality. We identify a phenomenon called *dual noise* to explain this stalling behavior and present an analysis of the desirable properties of dual solutions in this context. The proposed deflected subgradient approach has been embedded within the crew pairing solver at United Airlines and tested using historical data sets. Our computational experience suggests a strong correlation between the dual noise phenomenon and the quality of the final solution produced, as well as with the accompanying algorithmic performance. Although we observed that our deflected subgradient scheme yielded an average speed-up factor of 10 for the column generation scheme over the commercial solver, the average reduction in the optimality gap over the same number of iterations was better by a factor of 26, along with an average reduction in the dual noise by a factor of 30. The results from the column generation implementation suggest that significant benefits can be obtained by using the deflected subgradient-based scheme instead of a black-box-type or standard solver approach to solve the intermediate linear programs that arise within the column generation scheme.

*Key words*: airline crew planning; set-partitioning problems; deflected subgradient approach; dual noise; variable target value method
*History*: Accepted by William Cook, former Area Editor for Design and Analysis of Algorithms; received February 2006; revised April 2007; accepted November 2007. Published online in *Articles in Advance* May 15, 2008.



## 1. Introduction

Consider the following discrete optimization problem, which represents a set-partitioning problem with side constraints (SPP-SC) (see Nemhauser and Wolsey 1999):

$$\begin{aligned} \textbf{SPP-SC:}\quad & \text{Minimize} && c^T x \\ & \text{subject to} && Ax = e, \\ & && Dx \leq d, \\ & && x \in \mathrm{B}^n. \end{aligned} \tag{1}$$

Here, the variables $x \in \mathrm{B}^n$ represent binary decision variables, with an associated cost vector $c$. The first set of equality constraints in problem SPP-SC represent set-partitioning constraints with the coefficients $a_{ij}$ of the $m \times n$ matrix $A$ being either zero or one, depending on whether variable $x_j$ participates in constraint $i$ or not, and where $e$ is a vector of ones. The second set of some $r$ inequality restrictions are side constraints that arise out of certain designated goals pertaining to alternative multiobjective requirements, beyond the cost-minimization objective of SPP-SC.

A typical airline planning instance of SPP-SC is the multiobjective airline crew pairing optimization problem, where the columns $A_j$ of $A$ are generated to satisfy delineated crew pairing legality restrictions. More specifically, this problem seeks an optimal assignment of crew pairings to different flight segments, subject to various contractual and FAA-mandated crew scheduling regulations. In practice, airlines are interested in finding a (near-optimal) solution that would satisfy certain utilization, manpower availability, and quality-of-work-life (QWL) requirements such as schedule regularity, in addition to minimizing crew costs that include hotel, per diem, and traditional metrics such as flight-time-credit (FTC), and a variety of other segment and duty-based penalties and incentives. Vance et al. (1997) provide more details on these crew pairing costs, and Klabjan et al. (2001, 2002) consider such crew pairing problems with additional regularity and time-window restrictions, as well as include some aspects related to the aircraft-routing problem.

We will restrict the scope of our study in this paper to the pure set-partitioning problem SPP, whose structure poses the main challenge in solving large-scale instances to near optimality. The current analysis can be readily extended to deal with the more restricted and realistic model SPP-SC. In practice, to maintain feasibility, we add nonnegative slack variables $s$, along with a corresponding large positive undercoverage penalty $P$ to the objective function, where $P$ is typically chosen to be considerably larger than the cost of any legal pairing that can be generated for the given problem. This results in the following augmented formulation of SPP:

$$\textbf{SPP}: \quad \begin{aligned} &\text{Minimize} && c^T x + Ps \\ &\text{subject to} && Ax + s = e, \\ & && s \geq 0, \quad x \in \mathrm{B}^n. \end{aligned} \qquad (2)$$

We now present a brief summary of relevant crew pairing concepts used in crew scheduling applications. A *crew duty* represents an approximate day's worth of work and consists of a legal sequence of flight segments and flight connections. *Layovers* represent rest periods between successive duty schedules and these accrue both crew and hotel costs. A *crew pairing* is a sequence of crew duties, along with any intermediate layovers, which begins and ends at the same crew domicile. Duties and layovers have to satisfy a variety of FAA-mandated and contractual set of crew rules, such as "8-in-24" restrictions (no more than eight hours of flying time in any rolling 24-hour window), night-flying limitations, and maximum duty limits. For a more detailed discussion of crew duties, crew rules, connections, and layovers, we refer the reader to Vance et al. (1997) and the FAA website (http://www.faa.gov).

Typically, the universe of crew pairings is huge and may run in the order of trillions for even modest problem sizes. A standard solution approach for crew scheduling problems is to first use a column generation-based procedure to solve a linear programming relaxation of SPP and to then continue solving the underlying discrete optimization problem using branch-and-price techniques (Barnhart et al. 1998) along with follow-on fixing strategies (Vance et al. 1997). Makri and Klabjan (2004) describe some of the computational challenges associated with column generation approaches for solving practical crew scheduling applications. Also, the Carmen Systems group of Boeing, Inc. reports results that demonstrate their considerable success in solving massively sized railroad crew scheduling problems using column generation techniques (Kohl 2003).

Given the nature of the problem and various other business requirements, a sequential planning process has been traditionally followed, starting with a gross-day approximation where every flight is assumed to operate every day, and ending with a final solution that represents crew pairings that cover each of the flight segments by date. However, recent advances in computer hardware and new column generation algorithms and techniques have facilitated obviating traditional sequential and suboptimal crew planning approaches, and to directly solve the multiobjective crew scheduling problem in its entirety within a reasonable amount of computational time. Furthermore, this success has enabled the solution of large optimization problems that arise from integrated airline planning models that cross traditional boundaries between different airline business areas.

The problem instances we consider in this paper are from weekly and monthly crew pairing problems that pertain to United Airlines's flight schedules. These are large-scale SPP instances that directly address a typical week's abstracted problem or the scheduling problem for an entire month. The latter choice is made when a large number of dated exceptions are present in the schedule, which usually occurs during holiday seasons or during months that contain several scheduling changes. In fact, the intensely competitive nature of the airline industry has typically resulted in an increase in the frequency of such demand-driven dynamic scheduling changes to optimally match supply and demand. Typically, the number of flight segments in such data sets range from 5,000–15,000 after applying preprocessing techniques. As an additional complication, we also allow limited "dead-heading," wherein crew members are transported between stations, usually on commercial flights, allowing for more connection options in generating feasible pairings. In the problem instances we solve, the segments from the entire United Airlines's flight schedule for the month are available as potential candidates for dead-heading. Note that the deadhead segments for a crew pairing do not participate in the coverage constraints, but only affect the associated cost coefficient.

Another aspect of problem SPP in practice is that the size of the problem alone does not determine the difficulty of the problem. The resulting nonlinearity and structure of crew rules, the cost-objective structure, the trade-off between competing objectives, the sparsity of the flight segment network, peaking effects due to the hub-and-spoke airline structure, among other problem aspects, can greatly influence the fractionality of the relaxations, run times, and storage requirements.

### 1.1. Motivation for Our Study

In the column generation (CG) process for solving such large-scale crew pairing optimization problems, we frequently observed the phenomenon of stalling that practically results in the early termination of the CG scheme, without finding even a feasible solution, when used in conjunction with a commercial barrier or simplex-based linear programming (LP) solver. The stalling typically originates early in the solution procedure, when the LP solution after the first few CG iterations turns out to be integral or nearly integral. After such an iteration, due to the inherent degeneracy in the problem, the LP objective does not improve, even though the subproblem quite easily generates tens of thousands of negative reduced-cost columns. There have been several papers published in the literature that deal with schemes to accelerate CG techniques for SPP instances, but we have not found any work that analyzes the properties of the optimal dual solution in this context. Among related research, Kohl (2003) presents computational results that demonstrate the efficacy of using subgradient-based approaches in commercial crew scheduling applications. Hu and Johnson (1999) address stalling and convergence issues with their CG scheme for airline crew pairing problems and propose an enhanced subproblem approach that seeks to generate an improved dual solution at every iteration. Merle et al. (1999) propose a CG stabilization scheme to overcome the slow convergence effects of Kelley's (1960) cutting-plane method in the presence of degeneracy by including perturbation variables within the master program, in concert with a convergent penalty-based scheme to control the level of perturbation. Barnes et al. (2002) solve an auxiliary nonlinear least squares problem in their CG approach, leading to a formulation that is impervious to degeneracy. Barahona and Anbil (2000) employ the volume algorithm heuristic to quickly find approximate primal and dual solutions to the master program, and report a significant improvement in computational efficiency over prior methods. Carvalho (2005) derives a family of valid dual cuts that can accelerate CG and applies this method in the context of the cutting-stock problem. Elhallaoui et al. (2005) suggest a dynamic constraint aggregation method that reduces the number of set-partitioning constraints and present results in the context of urban mass-transit scheduling. Amor et al. (2006) attempt to generate deep dual-optimal cuts for improving the effectiveness of their CG procedure, and test this approach on the cutting-stock problem.

Another practical limitation with subgradient-based schemes proposed in the literature is the number of algorithmic parameters that need to be fine-tuned for the sake of an industrial implementation. Typically, the maintenance and updating of such parameter-intensive methods diminishes over the years, leading to a drop in solution quality, which often leads practitioners to revert to suboptimal schemes based on off-the-shelf LP solvers.

The main contributions of this work are to provide insights into the desirable characteristics of a dual solution that can overcome the observed stalling problem alluded above, and to develop an efficient deflected subgradient-based optimization strategy having relatively few parameters that can reasonably achieve such a dual solution in practice.

The remainder of this paper is organized as follows. Section 2 presents the overall CG scheme and discusses the stalling phenomenon in greater detail. Section 3 describes a procedure for solving the master program. Sample computational results are presented in §4, and §5 concludes the paper with a summary and directions for future research.

## 2. Column Generation Scheme

In lieu of solving problem SPP directly, we first solve a continuous relaxation, **R-SPP**, of the problem using a CG technique (see Nemhauser and Woolsey 1999) and then use a branch-and-price heuristic technique to find an acceptable integer solution to problem SPP. In this CG approach for solving R-SPP, the master problem, RMP, is a linear program that is a restriction of R-SPP, where only a subset of up to some $n' < n$ columns is included as shown below:

$$\begin{aligned}
\textbf{RMP}: \quad & \text{Minimize} \quad \sum_{j=1}^{n'} c_j x_j + P \sum_{i=1}^{m} s_i \\
& \text{subject to} \quad \sum_{j=1}^{n'} a_{ij} x_j + s_i = 1 \quad \forall\, i = 1, \ldots, m, \\
& \qquad x_j \geq 0 \quad \forall\, j = 1, \ldots, n', \\
& \qquad s_i \geq 0 \quad \forall\, i = 1, \ldots, m.
\end{aligned} \tag{3}$$

At any CG iteration $l$, RMP is solved to produce an optimal (or near-optimal) dual solution $\pi^l$. A pricing subproblem, designated SP($\pi^l$), is then solved that attempts to generate up to some $p \geq 1$ pairings that have the lowest reduced costs, $(c_j - (\pi^l)^{\mathrm{T}} A_j)$, where each column $A_j$ of $A$ satisfies all the mandated crew pairing legality requirements. If no (or few) such negative reduced columns exist, the CG procedure terminates. Otherwise, the new columns are added to the restricted master program, and RMP is reoptimized to find a new dual vector. The model RMP that results at the end of the CG procedure is chosen as the root node problem for solving the discrete optimization problem SPP. A branching scheme that guarantees convergence for 0–1 set-partitioning problems is the "follow-on" fixing technique based on a modified Ryan-Foster's rule (Vance et al. 1997). A branch-and-price framework is typically used to solve SPP, where the pricing module again utilizes the CG scheme to find good integer solutions to SPP. The overall scheme of the CG procedure is described below.

*Step* 1. Set $l = 1$, and $s_i = 1$ and $\pi_i = P\ \forall\, i = 1, \ldots, m$.

*Step* 2. Solve the subproblem SP($\pi^l$) to generate up to some $p \geq 1$ negative reduced cost columns and add these to RMP. If fewer than some $p_{\min}$ such columns are generated, terminate the procedure; otherwise, proceed to Step 3.

*Step* 3. Set $l \leftarrow l + 1$. Solve RMP, possibly only to near optimality, to generate a good quality dual solution $\pi^l$. If $\|\pi^l - \pi^{l-1}\|/m < \varepsilon_\pi$, for some tolerance $\varepsilon_\pi > 0$, or if $l = l_{\max}$, a specified maximum number of iterations, terminate the CG procedure, and if necessary, re-solve RMP to optimality to find an optimal primal-dual solution using any suitable LP solver, and enter the branch-and-price routine. Otherwise, go to Step 2.

### 2.1. Dual Noise Phenomenon in Column Generation

Consider a large instance of SPP in the absence of the dead-heading option. In this case, every feasible pairing will have to cover at least two segments. Now, consider a CG scheme that produces an integer-feasible LP solution $\bar{x}$ after a few iterations. Let us assume that all the slack variables take values of zero in this solution, or equivalently, assume that we have preprocessed the input data set to eliminate all uncoverable flight segments from consideration to ensure the existence of such an $\bar{x}$ for the resulting SPP instance. Such an integral LP solution to this problem is also an extreme point LP solution having at most $\lfloor m/2 \rfloor$ basic variables at unity (because each $A_j$-column has at least two nonzero entries), and therefore, at least $\lceil m/2 \rceil$ basic variables are equal to zero. This situation represents a highly degenerate extreme point, and we shall show that the quality of the corresponding dual-optimal solution and the choice of the subproblem methodology for generating columns can greatly inhibit any improvement in the LP objective in subsequent iterations of the CG scheme.

Toward this end, let $J$ represent the index set of the unit integer-valued variables in an optimal solution to Problem RMP, where $|J| \leq \lfloor m/2 \rfloor$. Let us examine the characteristics of the dual solution under this situation by analyzing RD, the dual to problem RMP:

$$\begin{aligned}
\textbf{RD}:\quad & \text{Maximize} \ \sum_{i=1}^{m} \pi_i \\
& \text{subject to} \ \sum_{i=1}^{m} a_{ij}\pi_i \leq c_j \quad \forall\, j = 1, \ldots, n, \\
& \pi_i \leq P \quad \forall\, i = 1, \ldots, m, \\
& \pi \text{ unrestricted.}
\end{aligned} \tag{4}$$

Note that RD is a relaxation of the dual problem associated with the unrestricted master program that includes all possible columns because only a subset of all possible dual constraints are considered here. Let us define the set $S_j$ to be the index set of the rows of SPP that contain the variable $x_j$. Because $\bar{x}$ is an optimal solution to RMP, any dual-optimal solution $\pi^*$ to RD is characterized by the following linear system, denoted $\Pi^*(J)$:

$$\begin{aligned}
& \sum_{i \in S_j} \pi_i^* = c_j \quad \forall\, j \in J, \\
& \sum_{i \in S_j} \pi_i^* \leq c_j \quad \forall\, j \notin J, \\
& \pi_i^* \leq P \qquad \forall\, i = 1, \ldots, m.
\end{aligned} \tag{5}$$

As an illustrative example, consider the following four-segment SPP that arises from a gross-day crew pairing problem, where the rows are covered after the first column generation iteration. Suppose that this iteration adds exactly two pairings containing the first two and last two segments, respectively, with identical high costs of $P/2$. Therefore, the set of dual-optimal solutions to RMP at this iteration satisfies

$$\pi_1^* + \pi_2^* = P/2, \tag{6a}$$

$$\pi_3^* + \pi_4^* = P/2, \tag{6b}$$

$$\pi_i^* \leq P, \quad i = 1, \ldots, 4. \tag{6c}$$

One possible solution that satisfies these conditions is given by

$$\pi_1^* = \pi_3^* = P/4 + \sigma, \tag{6d}$$

$$\pi_2^* = \pi_4^* = P/4 - \sigma, \tag{6e}$$

where

$$-\frac{3P}{4} \leq \sigma \leq \frac{3P}{4}. \tag{6f}$$

In particular, setting $\sigma = 3P/4$ gives us $\pi_1^* = \pi_3^* = P$, and $\pi_2^* = \pi_4^* = -P/2$, which is an extreme point dual-optimal solution. Suppose now that we proceed to the next step of solving the subproblem with this dual solution and generate two new pairings $x_3$ and $x_4$, with the first pairing having a cost of $c_3$ and covering segments 1 and 3, and the second pairing having a cost of $c_4$ and covering segments 2 and 4. Hence, $x_3$ will have a reduced cost of $c_3 - 2P$, and $x_4$ will have a reduced cost of $c_4 + P$. This implies that any pairing of the type $x_3$ will have a negative reduced cost and will enter the matrix as long as $c_3 < 2P$, a large number, whereas $x_4$ will not enter the matrix unless $c_4$ is highly negative ($< -P$), which is almost unrealizable in practice. In particular, $c_3 = P$ and $c_4 = 0$ will result in the interesting situation where $x_3$ enters the matrix, leading to no improvement in the objective (although it cuts off the current optimal dual solution). Let us assume that there exists an additional legal pairing $x_5$ having a cost of zero that covers all the segments in the sequence 2, 4, 1, and 3. The partial path of $x_5$ containing its first two segments resembles $x_4$ and is a candidate that is likely to be eliminated by the heuristic CG subproblem schemes that employ labeling-based resource-constrained shortest-path methods in practice (see Desrosiers et al. 1995), because after the stage of generating the first two segments, it will have accumulated an unfavorably high reduced-cost estimate.

Figure 1 illustrates the highly nonlinear fluctuation in reduced-cost estimates of partial pairings observed within the subproblem heuristic associated with a historical United Airlines weekly crew scheduling data set. The figure shows the variation in the reduced-cost estimate after each label extension that adds another flight segment to the pairing. Thicker lines delineate pairings that exhibit extreme variations. Each of the pairings shown in the figure successfully priced out with a negative reduced cost at the end of the labeling heuristic.

This example illustrates the potential pitfalls of a black-box-type application of a commercial LP solver within an implementation of such a dynamic CG procedure. We term this phenomenon *dual noise*, whereby certain inordinately high dual component magnitudes dominate the reduced-cost estimate of a partial pairing, rendering the original objective cost component insignificant in comparison and resulting in poor columns being generated. Note also that in the presence of significantly high dual noise, the subproblem scheme would need to process a significantly higher number of partial pairings to add any meaningful subset of columns to the matrix that might lead to an improvement in the objective value. Furthermore, as explained in the sequel, the addition of columns driven by dual noise tends to generate weak inequalities in the dual space, which results in a near-stationary dual solution as well.

For the large crew scheduling problems we tackle here, the cardinality of $J$ is typically close to $m/5$, leading to substantially higher degeneracy and dual noise. In such instances that we solved for United Airlines, the black-box solver approach had to be abandoned due to a persistent stalling of the objective value. Furthermore, in a typical dynamic CG implementation, whenever we reach the preset column storage limit, columns having high reduced costs are replaced by those that are newly generated. Doing so creates further problems because the presence of dual noise clouds the quality of columns, rendering such column replacement schemes ineffective.

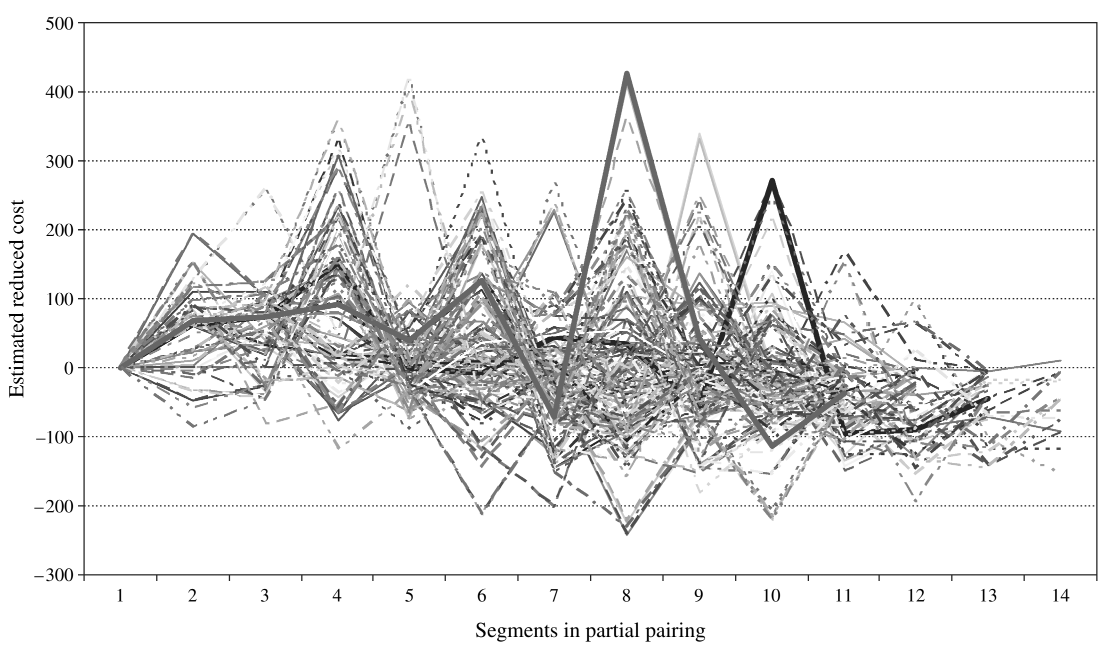


**Figure 1 Progress in the Reduced-Cost Estimate of Partial Pairings Within the Subproblem Heuristic**

In practice, the reduction of dual noise allows us to significantly reduce the preset storage limit, without any adverse impact on the quality of the solution to SPP. In general, interior point dual solutions seem to better suit column generation schemes in the context of SPP when compared with extreme point solutions obtained via the simplex method. Furthermore, dual solutions obtained via subgradient optimization procedures have yielded better results in large SPP instances when compared with those obtained using a commercial optimization software package. A similar conclusion has been independently observed in other commercial crew pairing applications on the basis of several empirical tests (Hjorring 2004) and in vehicle-routing and staffing problems (Gunluk et al. 2004).

From the previous example, it is clear that not all dual-optimal solutions are equally desirable. For example, a choice of $\sigma = 0$ in the foregoing illustration would have resulted in all components of the dual solution taking a positive value of $P/4$. This choice significantly improves the possibility of generating better columns via the subsequent subproblem and, consequently, achieving a nonzero improvement in the LP objective. Interestingly, this dual solution also has the smallest Euclidean norm among all alternative dual-optimal solutions. (For brevity, "norm" and $\|\cdot\|$ will denote the Euclidean norm throughout the paper.) A key observation is that, in practice, we would like to work with dual solutions that have norms of low magnitude. It is our experience that minimal norm dual solutions lead to reduced dual noise. This can be intuitively explained as follows. We would like to improve the effectiveness of the subproblem heuristic by actively seeking dual solutions that do not dominate the reduced-cost estimates of partial pairings, and thereby limit the occurrence of large partial reduced-cost variations over the labeling steps. Moreover, we would like to generate columns that lead to deep cuts in the dual space. Denoting $\Pi^*$ as the set of optimal dual solutions to a current RMP, note that we would ultimately like to find a $\pi \in \Pi^*$ such that for some $j$, we obtain a negative reduced cost $c_j - \pi^T A_j = c_j - \sum_{i\in S_j} \pi_i$. Generating such a negative reduced-cost column and adding $\sum_{i\in S_j} \pi_i \le c_j$ to the dual problem would yield a dual cut. We would like to cut off as much of the current $\Pi^*$ as possible, and therefore, we would like to find an enterable column index $j$ for which $\max_{\pi\in\Pi^*}\{c_j - \sum_{i\in S_j} \pi_i\} < 0$; i.e., $\{c_j - \min_{\pi\in\Pi^*} \sum_{i\in S_j} \pi_i\} < 0$. Hence, noting that the components of $\pi$ tend to be nonnegative at optimality due to the covering ($\ge$) tendency of the set-partitioning constraints, if we use a $\pi \in \Pi^*$ having relatively small components that yet generates a negative reduced-cost column, then this will likely yield a strong cut giving a dual ascent. Noting that the dual objective function is to maximize $e^T\pi$, we could therefore choose to find an optimal dual solution that among alternative optima, minimizes $\|\pi\|^2$; i.e., we could consider solving

$$\textbf{MDN}: \quad \begin{aligned} &\text{Minimize} \quad \|\pi\|^2 \\ &\text{subject to} \quad \pi \in \Pi^*. \end{aligned} \tag{7}$$

Problem MDN may be burdensome to use in practice. Instead, we can find a near-optimal solution to the following dual quadratic program, DQP, which is the dual to RMP that is augmented by a quadratic term, where $\omega$ is a small positive penalty on the dual norm:

$$\textbf{DQP}: \quad \begin{aligned} \text{Maximize} \quad & e^T\pi - \frac{\omega}{2}\|\pi\|^2 \\ \text{subject to} \quad & A^T\pi \le c, \\ & \pi \le Pe, \\ & \pi \text{ unrestricted.} \end{aligned} \tag{8}$$

We can also thereby measure the dual noise for a selected dual solution $\pi^*$ using the *dual noise factor* $\eta$, given by

$$\eta = \frac{\|\pi^*\|}{\|\pi^*_{\min}\|}, \tag{9}$$

where $\pi^*_{\min}$ is a dual-optimal solution having the smallest Euclidean norm. We expect that a subproblem heuristic that uses any dual-optimal (or near-optimal) solution having a relatively small value of $\eta$ to generate negative reduced-cost columns will result in relatively strong cuts in the dual space.

Note that problem DQP is a concave QP maximization problem. In lieu of problem DQP, we can equivalently examine its Dorn dual formulation to obtain the following QP extension of RMP (see Bazaraa et al. 2006, for example):

$$\textbf{MQP}: \quad \begin{aligned} \text{Minimize} \quad & c^Tx + Pe^Ts + \frac{\omega}{2}z^Tz \\ \text{subject to} \quad & Ax + s + \omega z = e, \\ & x, s \ge 0, \quad z \text{ unrestricted.} \end{aligned} \tag{10}$$

Observe that problem RMP can be viewed as a special case of MQP with $\omega = 0$. Solving MQP using some positive $\omega$ is equivalent to adding a perturbation $\omega z$ to the constraints in (3) and including a quadratic penalty term in the objective that limits such a perturbation. This form is related to ideas that attempt to stabilize column generation. Merle et al. (1999) propose an alternative perturbation scheme and penalize the rectilinear norm of the perturbation vector. Likewise, Barnes et al. (2002) propose a nonnegative least squares (NNLS) problem that aims to minimize a perturbation measure over the set of active constraints. Also, interestingly, Gunluk et al. (2004) report an improvement in the performance of their CG procedure for a vehicle-routing problem by using a heuristic discounting scheme that effectively reduces the magnitude of the optimal dual vector.

Here, we choose to solve the Lagrangian duals to RMP and MQP using a new deflected subgradient optimization method that is presented below. Note that the optimality conditions for MQP with respect to the $z$-variable column requires that $z = \pi$, so that similar to the case of RMP, we obtain a Lagrangian dual subproblem that involves only the $(x, s)$-variables. The deflected subgradient scheme that we propose not only yields near-optimal dual solutions to RMP and MQP having a small Euclidean norm but is also computationally inexpensive when compared with a commercial LP solver applied to solving problem RMP. Moreover, on an average, we obtain a reduction of more than 75% in dual norm values when compared with this commercial solver as we shall see in §4. We combine this deflected subgradient method with an efficient implementation of the resource-constrained shortest-path approach for solving the column generation subproblem, which manages dual noise in an efficient manner without significantly increasing the computational burden.

## 3. A Deflected Subgradient Method for Solving Linear Programs

In this section, we present some promising subgradient-based methods for solving the Lagrangian duals to problems MQP and RMP. These approaches are nondifferentiable optimization (NDO) methods that involve moving along deflected subgradient (DSG) directions to derive near-optimal, minimum-norm dual solutions to RMP and MQP in the context of column generation.

Baker and Sheasby (1999) have shown that by applying exponential smoothing to the subgradient vector, the convergence process can be accelerated. The volume algorithm (VA) of Barahona and Anbil (2000) is based on an efficient approximation scheme similar to Baker and Sheasby's (1999) method of exponential smoothing, and uses moving averages to generate an estimate of the primal vector along with a heuristic dual solution. However, the VA in its original form does not guarantee primal or dual convergence, but could be modified to obtain dual convergence. For example, Sherali and Lim (2004) have enhanced the VA concept by viewing it as a special case of the deflected subgradient approach in the dual space, and have accordingly embedded the VA dual update scheme in the convergent variable-target value method (VTVM) developed by Sherali et al. (2000) to obtain encouraging results. In §4, we provide some comparative computational results using this scheme as well.

### 3.1. Dual-Norm-Based Deflected Subgradient Method (DSG)

Consider the following linear program LP, which represents RMP in convenient form, and its corresponding Lagrangian Dual LD:

$$\begin{aligned} \textbf{LP}: \quad & \text{Minimize} \quad c^T x \\ & \text{subject to} \quad Ax = b, \\ & \qquad\qquad\quad x \in \bar{X}, \end{aligned} \tag{11}$$

where

$$\bar{X} = \{x: 0 \le x \le e\}.$$

$$\textbf{LD}: \quad \text{Maximize}\{\theta(\pi): \pi \text{ unrestricted}\}, \tag{12}$$

where $\theta(\pi)$ is given by the optimal value to the Lagrangian subproblem:

$$\textbf{LS}(\pi): \quad \text{Minimize}\{c^T x + \pi^T(b - Ax): x \in \bar{X}\}. \tag{13}$$

The method we propose is an implicit variable target value approach that can be used to quickly generate good-quality solutions to LD in practice. Our goal is to construct a sequence of incumbent dual vectors by using the product of the Euclidean norms of the incumbent dual vector and the current search direction to assess an implicit target value for determining the step size for generating the next dual vector, where the search direction is derived using a suitable convex combination of the current subgradient and the previous direction. We provide some motivation for this choice of step length below.

**3.1.1. Choice of Step Length.** Given a nondifferentiable concave function $\theta(\pi)$, and a search direction vector $d$ belonging to the subdifferential at $\pi$, we obtain the following inequalities by the concavity property of $\theta$ and using the Cauchy-Schwarz inequality, where $\pi^*$ solves LD:

$$\theta(\pi^*) - \theta(\pi) \le (\pi^* - \pi)^T d \le (\|\pi^*\| + \|\pi\|)\|d\|. \tag{14}$$

For any dual solution $\pi$, let us define an *implicit target value difference*

$$\Delta T = T - \theta(\pi), \tag{15}$$

where $T$ is an *implicit target value* at any particular iteration, with the restriction that it cannot be lesser than some

$$T_{\min} = \theta(\pi) + \gamma \varepsilon_T \|d\|^2 \tag{16}$$

for a given $\varepsilon_T > 0$. Let the implicit target value be such that $\Delta T = \gamma\|\pi\|\|d\|$, so that

$$\begin{aligned} & T \equiv \theta(\pi) + \gamma\|\pi\|\|d\| \quad \text{and} \\ & T \ge T_{\min} \;\Rightarrow\; \gamma\|\pi\| \ge 2\varepsilon_T\|d\|, \end{aligned} \tag{17}$$

where $\gamma$ is a factor prescribed below depending on two possibilities that can arise.

*Case* 1. $\|\pi^*\| \le \|\bar{\pi}\|$. In this case, (14) yields $\theta(\pi^*) - \theta(\bar{\pi}) \le 2\|\bar{\pi}\|\|d\|$, and noting (17), we therefore obtain an implicit target value $T$ that is an upper bound on the optimal objective value by selecting $\gamma \ge 2$.

*Case* 2. $\|\bar{\pi}\| < \|\pi^*\|$. This implies from (14) that $\theta(\pi^*) - \theta(\bar{\pi}) \le 2\|\pi^*\|\|d\|$. Depending on the current dual norm, two subcases arise.

*Case* 2a. $2\|\bar{\pi}\|\|d\| \le \theta(\pi^*) - \theta(\bar{\pi}) \le 2\|\pi^*\|\|d\|$. In this situation, by (17) a choice of $\gamma \le 2$ yields an implicit target value $T$ that is a lower bound on the optimal Lagrangian dual value.

*Case* 2b. $\theta(\pi^*) - \theta(\bar{\pi}) \le 2\|\bar{\pi}\|\|d\| \le 2\|\pi^*\|\|d\|$. This situation is similar to Case 1, so again, a choice of $\gamma \ge 2$ results in an implicit target value $T$ that is an upper bound on the optimal objective.

Motivated by these cases, we select a value of $\gamma = 2$ to construct our implicit target value as prompted by (17) according to

$$T \equiv \theta(\bar{\pi}) + 2\|\bar{\pi}\|\|d_t\| \text{ at iteration } t, \quad \text{with } T_{\min} \equiv \theta(\bar{\pi}) + 2\varepsilon_T\|d_t\|^2, \tag{18}$$

where $\bar{\pi}$ and $d_t$ (to be specified) are, respectively, the incumbent dual vector and the search direction at iteration $t$. The prescribed dual update scheme is then given by

$$\pi_{t+1} = \bar{\pi} + \frac{\beta[T - \theta(\bar{\pi})]}{\|d_t\|^2} d_t = \bar{\pi} + \frac{2\beta\|\bar{\pi}\|\|d_t\|}{\|d_t\|^2} d_t = \bar{\pi} + \frac{2\beta\|\bar{\pi}\|}{\|d_t\|} d_t, \tag{19}$$

where $0 < \beta_{\min} \le \beta \le 1$ is a suitable step-length parameter. This yields the following bounds on the resulting norm of the dual solution by the Cauchy-Schwartz inequality:

$$(1 - 2\beta) \le \frac{\|\pi_{t+1}\|}{\|\bar{\pi}\|} \le (1 + 2\beta) \Rightarrow -2\beta \le \frac{\|\pi_{t+1}\| - \|\bar{\pi}\|}{\|\bar{\pi}\|} \le 2\beta. \tag{20}$$

Starting with $\pi_1 = 0$, $t = 1$, and $d_1 = g_1$, the dual update (19) results in a sequence of dual solutions that attempts to improve the Lagrangian dual objective function $\theta(\cdot)$ by adopting a prescribed step of length $2\beta\|\bar{\pi}\|$ along the normalized DSG direction $d_t/\|d_t\|$ at iteration $t$. Note that whenever $T$ is less than $T_{\min}$ as defined in (18), we use $T_{\min}$ in lieu of $T$ in (19), giving the dual update scheme

$$\pi_{t+1} = \bar{\pi} + 2\beta\varepsilon_T d_t. \tag{21}$$

As far as the direction $d_t$ is concerned, several deflection schemes have been prescribed in the literature (see, for example, the discussion in Sherali and Choi 1996). Any of these schemes could be used in the above method. We shall adopt an exponential smoothing scheme that appears to empirically accelerate dual convergence, with the added simplification of simply choosing the step-length parameter $\beta$ itself as the coefficient for the smoothing scheme. Under this approach, the direction $d_t$ is derived as a convex combination of the current subgradient direction and the previous search direction vector, as given below:

$$d_t = \beta g_t + (1 - \beta) d_{t-1}, \tag{22}$$

where $g_t$ is a subgradient of $\theta$ at $\pi_t$, which is obtained by solving the Lagrangian subproblem given by (13), and $\beta$ is the current step-length parameter. Letting $x_t$ be an optimal solution to LS($\pi_t$), we can use $g_t = b - Ax_t$. Observe that $d_t$ is essentially a deflected subgradient direction given by $(1/\beta)d_t = g_t + \psi d_{t-1}$, where $\psi \equiv (1 - \beta)/\beta$. The resulting expression is of a similar general type to the alternative specific procedures used in Baker and Sheasby (1999) and Barahona and Anbil (2000).

Note that if we use any convergent variable-target-based scheme (such as that in the VTVM method of Sherali et al. 2000, for example), the resultant deflected subgradient methodology based on (22) satisfies the requirement for convergence to a dual-optimal solution (Sherali and Lim 2004). Using the self-correcting target scheme (18) in concert with (19), (21), and (22) reduces the number of algorithmic parameters by at least two when compared with the VTVM scheme.

To summarize the proposed deflected subgradient method for optimizing LD, we start with an initial dual incumbent $\bar{\pi} \equiv 0$ as the iterate $\pi_t$ at $t = 1$, and define $\bar{\theta} \equiv \theta(\bar{\pi})$, set $\varepsilon_T = 1$, and use $d_1 \equiv g_1$ as the search direction. The dual solution is then updated at any iteration $t$ by taking a positive step length proportional to the incumbent dual solution's norm in the normalized DSG direction as stated below, where we have used (18), (19), and (21) to obtain (23a):

$$\pi_{t+1} = \begin{cases} \bar{\pi} + 2\beta \dfrac{\|\bar{\pi}\|}{\|d_t\|} d_t & \text{if } \|\bar{\pi}\| \ge \varepsilon_T \|d_t\| \text{ and} \\ \bar{\pi} + 2\beta\varepsilon_T d_t & \text{otherwise.} \end{cases} \tag{23a}$$

$$\theta_{t+1} = \min_{x \in \bar{X}} \{c^T x + \pi_{t+1}^T (b - Ax)\} = c^T x_{t+1} + \pi_{t+1}^T g_{t+1}, \quad \text{where } g_{t+1} \equiv b - Ax_{t+1}. \tag{23b}$$

$$\text{If } \theta_{t+1} > \bar{\theta}, \quad \text{then set } \bar{\pi} = \pi_{t+1} \text{ and } \bar{\theta} = \theta_{t+1}. \tag{23c}$$

$$\text{Let } d_{t+1} = \beta g_{t+1} + (1 - \beta) d_t, \text{ and reiterate.} \tag{23d}$$

In practice, we also use the projected quadratic-fit line search during the first 1,000 iterations to accelerate convergence, as described in Lim and Sherali

(2006). Note that the dual update in (23a) requires a step from the incumbent dual vector, unlike most other subgradient optimization schemes that adopt a step length along the (deflected) subgradient direction from the most recent dual vector. We observed that the latter scheme, while being competitive in terms of the final solution gap, tends to produce relatively larger dual norms in practice. Furthermore, (20) specifies that the bounds on the norm of the next dual iterate in the DSG scheme are proportional to the norm of the dual vector from which we step. Toward the aim of obtaining relatively smaller dual norms, especially when solving Problem MQP, we can modify (23a) in light of (20) by stepping from the most recent dual vector $\pi_t$ (and also defining $T$ and $T_{\min}$ with $\bar{\pi} \leftarrow \pi_t$ in (18)) whenever $\|\pi_t\| < \|\bar{\pi}\|$. This leads to the alternative update scheme:

$$\pi_{t+1} = \begin{cases} \hat{\pi} + 2\beta \dfrac{\|\hat{\pi}\|}{\|d_t\|} d_t & \text{if } \|\hat{\pi}\| \geq \varepsilon_T \|d_t\| \quad \text{and} \\ \hat{\pi} + 2\beta\varepsilon_T d_t & \text{otherwise,} \end{cases} \tag{23e}$$

where

$$\hat{\pi} = \begin{cases} \pi_t & \text{if } \|\pi_t\| < \|\bar{\pi}\| \quad \text{and} \\ \bar{\pi} & \text{otherwise.} \end{cases} \tag{23f}$$

In our proposed procedure using either (23a) or (23e) and (23f), the value of $\beta$ is initially set to 1.0, and is periodically halved whenever the incumbent objective improvement falls below a threshold level as shown below. Specifically, let $\varepsilon_\theta = 0.1$ be a threshold value for the ascent in incumbent objective achieved over the most recent $\tau_{\min} = 10$ iterations, below which the value of $\beta$ is halved. The value of $\tau$ that represents the next iteration to check for updating $\beta$ is initialized at $\tau_{\min} + 1 = 11$, the previous incumbent reference value is initialized at $\hat{\theta} = \bar{\theta}$, and the minimum permissible value for $\beta$ is set at $\beta_{\min} = 10^{-4}$. Then, the scheme for managing $\beta$ operates as follows, where $\bar{\theta}$ represents the current incumbent objective value at iteration $t$. If $t = \tau$, then do: If $\bar{\theta} - \hat{\theta} < \varepsilon_\theta$, then set $\beta \leftarrow \max\{\beta_{\min}, \beta/2\}$; else, retain $\beta$. Also, reset $\hat{\theta} \leftarrow \bar{\theta}$ and increment $\tau \leftarrow t + \tau_{\min}$.

The DSG procedure can be terminated when $\max_{i=1,\ldots,m} |d_{ti}| < \varepsilon_d$, where $\varepsilon_d$ is some small positive tolerance or if we reach a specified maximum iteration limit.

In lieu of directly solving the Lagrangian dual LD formulated for the linear program RMP, we can alternatively solve a similar Lagrangian dual formulated for MQP using the same approach, along with the modifications specified by (23e) and (23f), without any significant increase in computational effort required per iteration. We present some computational experience on applying the prescribed algorithmic strategies to both LD and the Lagrangian dual to MQP in the next section.

Convergence to an optimal solution can be induced by using the following two-phase approach. In phase I, the implicit target value DSG heuristic described above can be run for up to $t^{\mathrm{I}}_{\max}$ iterations. We can then switch over to the convergent VTVM-based target value algorithm of Lim and Sherali (2006) for phase II, using the same deflected subgradient-based direction. In this process, the implicit target value obtained at the end of phase I can be gainfully employed as the initial target value for phase II. Given the heuristic nature of phase I, and the demonstrated effectiveness of the Lim and Sherali (2006) procedure, this phase II augmentation serves as a useful practical safeguard in attaining good quality dual solutions via the overall two-phase scheme, besides simply ensuring theoretical convergence. An alternative approach would be to use the dual estimate obtained at the end of phase I to initialize the dual simplex method in phase II, in lieu of using a convergent NDO procedure such as VTVM.

## 4. Computational Experience

In this section, we provide sample results on five large-scale crew pairing SPP test sets that are derived from historical schedules at United Airlines. We also embed the proposed DSG scheme for RMP and MQP within ACRUZER©, United's Parallel Crew Pairing Solver, to evaluate its performance within the CG framework described in §2 using five other real-life proprietary test cases selected from a variety of fleet types and different operational crew regulations. In addition, we compare the performance of DSG to that of the interior point barrier solver and the dual simplex solver in CPLEX 9.1. In this context, we report statistics on the percentage optimality gap after every 1,000 iterations of the DSG scheme and the CPU run time after 3,000 iterations, as well as the percentage ratio of the DSG method's final dual norm to the corresponding value obtained for the barrier solver of CPLEX 9.1. To focus purely on algorithmic performance, all DSG runs were made with no prior knowledge of the optimal solution, no preprocessing of the problem or exploitation of inherent network structures to reduce its size, and starting with a zero dual vector in each case. For our implementation, we have used a Linux-based Intel Pentium 4, 2.6 GHz computer having 1.5 GB of RAM, and with serial processing only. The CPLEX barrier solver was applied to solve the five SPP examples to optimality using default parameter settings and with no crossover to the simplex method. We also present results for the MQP-based approach and the convergent VTVM scheme.

Table 1 presents statistics for the computational time and the solution value obtained by solving the

**Table 1** Computational Results Using CPLEX-Barrier to Solve the LP Relaxation of SPP to Optimality

| Data set | Rows | Columns | Final objective | Dual norm | CPU time (sec.) |
|---|---|---|---|---|---|
| CREW1 | 10,683 | 102,400 | 393,378.04 | 186,889 | 141 |
| CREW2 | 4,673 | 59,255 | 64,777.34 | 83,350.4 | 167 |
| CREW3 | 15,360 | 163,373 | 533,734.24 | 658,642 | 401 |
| CREW4 | 15,511 | 69,989 | 483,587.12 | 72,319.1 | 201 |
| CREW5 | 2,296 | 20,334 | 942,110.23 | 1,935,920 | 17 |

LP relaxation of SPP to optimality using the barrier solver option of CPLEX 9.1. Table 2 displays the results for solving the same LP relaxation via its Lagrangian dual using a stand-alone version of the proposed DSG solver. We can observe from Table 2 that the DSG method found a dual solution to within 10% of the optimal objective value after no more than 3,000 iterations for all the data sets. Furthermore, the Euclidean norm of the DSG-generated dual vector is significantly smaller than that generated by CPLEX (where the former is expressed as a percentage of the latter in the penultimate column of Table 2), being about 6%–43% of the latter value in magnitude, and yielding an average reduction factor of 4.5 in the dual noise factor $\eta$ (see Equation (9)). Furthermore, as evident from Table 3, it is possible to generate smaller dual norms by solving the Lagrangian dual to MQP, where we have used a relatively small value of $\omega = 10^{-7}$, and applied (23e) and (23f) instead of (23a) for MQP. When compared with Table 2, we obtained a decrease in the optimality gap for the largest CREW3 instance but with some degradation in solution quality for the other four cases. The last column of Table 3 indicates a reduction in the magnitude of the dual norm in all cases, yielding an average decrease of 24% beyond the dual norms generated by solving LD. Although greater reductions in the magnitude of the dual norm may be possible by using larger values for $\omega$, using disproportionately large values of this parameter could lead to increasingly suboptimal solutions. We recommend solving the Lagrangian dual to Problem MQP whenever a significant reduction in dual noise is critical for avoiding stalling of the column generation scheme. We report on such examples later in Table 5.

**Table 2** Results for the DSG Solver Using an Iteration Limit of 3,000

| Data set | Gap-1,000 % | Gap-2,000 % | Gap-3,000 % | Dual norm % ratio | CPU time (sec.) for 3,000 iter. |
|---|---|---|---|---|---|
| CREW1 | 0.62 | 0.35 | 0.27 | 8.64 | 30 |
| CREW2 | 6.99 | 1.30 | 0.57 | 19.00 | 18 |
| CREW3 | 8.37 | 6.05 | 6.05 | 5.62 | 48 |
| CREW4 | 1.45 | 0.91 | 0.79 | 34.95 | 21 |
| CREW5 | 48.28 | 16.20 | 8.38 | 42.73 | 3 |

**Table 3** Results for the DSG Solver Applied to the Lagrangian Dual of Problem MQP

| Data set | Gap-3,000 % | MQP/LD dual norm % ratio | CPU time (sec.) for 3,000 iter. |
|---|---|---|---|
| CREW1 | 0.61 | 60.53 | 30 |
| CREW2 | 5.88 | 54.84 | 18 |
| CREW3 | 5.89 | 95.02 | 49 |
| CREW4 | 1.27 | 76.08 | 21 |
| CREW5 | 9.07 | 94.59 | 3 |

Next, we present in Table 4 the performance of the convergent VTVM method using the generalized Polyak-Kelley cut (GPKC) scheme, as described in Lim and Sherali (2006). The DSG method performs significantly better in terms of run time and solution quality for these crew planning instances, where the goal is to quickly obtain good dual solutions. In all our test cases, the phase I solution itself was within an acceptable tolerance. However, to provide a practical safeguard, and to ensure theoretical convergence to optimality (or near optimality), we recommend running the two-phase DSG-VTVM scheme or the DSG-dual simplex scheme described in §3.

Finally, Table 5 compares the performance of the ACRUZER© crew solver using CPLEX-barrier, CPLEX-dual simplex, and the DSG scheme applied to each of RMP and MQP. A relative complementarity tolerance of 0.1 was used for the barrier method, while the DSG scheme was run using a tolerance value of $\varepsilon_d = 0.05$, and with an iteration limit of 2,000. All approaches, except the barrier-based solver, used an advanced start to initialize any CG iteration, either using the most recent basis for the dual simplex method, or the most recent dual estimate for the DSG approach. We have used a relatively large value of $\omega = 10^{-4}$ for the MQP-based scheme within ACRUZER to aggressively seek dual solutions having small norm values and thereby analyze the resultant behavior of the DSG scheme under low dual noise conditions. The statistics presented in Table 5 are derived from the ACRUZER runs made on a representative set of large-scale SPP instances that were encountered in historical scheduling months. Although the data sets in this section are proprietary, a relatively simple way to generate synthetic SPP data sets is to extract flight schedule data from the Official

**Table 4** Results for the VTVM-GPKC Scheme

| Data set | Gap-3,000 % | CPU time (sec.) |
|---|---|---|
| CREW1 | 4.34 | 89 |
| CREW2 | 25.24 | 62 |
| CREW3 | 14.86 | 162 |
| CREW4 | 7.93 | 69 |
| CREW5 | 52.0 | 7 |

**Table 5 Comparative Results Within ACRUZER**

| Data set | Flight segments | CG iterations | RMP solver | Final LP objective value | Euclidean dual norm | CPU time (minutes) |
|---|---|---|---|---|---|---|
| PIL67 | 2,000 | 10 | BARRIER | 2,291,459 | 2,260,587 | 4 |
| | | | SIMPLEX | 2,513,306 | 2,383,343 | 14 |
| | | | DSG-RMP | 2,330,865 | 560,658 | 2 |
| | | | DSG-MQP | 2,311,708 | 317,958 | 2 |
| PIL37 | 5,000 | 20 | BARRIER | 17,665,559 | 3,308,198 | 298 |
| | | | SIMPLEX | 12,166,390 | 3,454,063 | 473 |
| | | | DSG-RMP | 11,444,565 | 2,115,149 | 14 |
| | | | DSG-MQP | 7,927,230 | 92,107 | 16 |
| FLTAT | 16,000 | 30 | BARRIER | 26,663,988 | 5,071,235 | 610 |
| | | | SIMPLEX | 30,629,052 | 5,286,123 | 644 |
| | | | DSG-RMP | 13,666,756 | 4,439,346 | 20 |
| | | | DSG-MQP | 10,652,293 | 399,359 | 67 |
| PIL20 | 10,000 | 40 | BARRIER | 272,271,820 | 5,686,905 | 516 |
| | | | SIMPLEX | 291,986,340 | 5,849,908 | 428 |
| | | | DSG-RMP | 99,237,572 | 4,387,686 | 39 |
| | | | DSG-MQP | 2,313,577 | 83,969 | 54 |
| STALL | 5,000 | 60 | BARRIER | 7,003,305 | 2,506,208 | 377 |
| | | | SIMPLEX | 3,257,468 | 2,752,465 | 358 |
| | | | DSG-RMP | 308,713 | 87,107 | 14 |
| | | | DSG-MQP | 439,265 | 89,151 | 26 |

Airline Guide website (http://www.oag.com) for any particular week or month to generate the rows and incorporate the FAA crew regulations as well as any relevant contract-specific rules, if available, within the subproblem procedure to generate realistic columns. The data sets are arranged in increasing order of difficulty and represent different aircraft fleet types. All examples except the FLTAT data set (which is based on the crew regulations for flight attendant scheduling), use pilot scheduling rules. Table 5 reports on the problem size in terms of the number of flight segments, the number of CG iterations performed within ACRUZER, the final (optimal) LP objective value achieved, the Euclidean dual norm obtained after the penultimate CG iteration (i.e., before reoptimizing the final RMP to optimality using CPLEX), and the CPU run time for the CG scheme for each of the four approaches. The final LP objective value reported in Column 5 corresponds to the optimal LP solution at the end of the CG procedure (computed by solving the LP relaxation of the final RMP using CPLEX). The CPU run time required to resolve the continuous relaxation of the master problem to optimality, as well as the run time within the subproblem heuristic across all iterations, are included in the CPU time reported in the final column. The last data set (STALL) illustrates a real-life example in which the CG procedure based on a commercial solver stalls and fails to reasonably converge after a large number of CG iterations.

Although the performance of the CPLEX-based solvers was somewhat competitive for the smallest data set (PIL67), we observed that as the SPP problem increased in size and complexity, the DSG-based approach increasingly dominated the former approach. The results from the last two data sets (PIL20 and STALL) indicate the drastic failure of the barrier- and simplex-based solvers in providing a practically useful set of columns at the end of the procedure. Furthermore, the black-box CG approach frequently had to be abandoned in practice for virtually all large data sets because the preset column storage limit was quickly reached and no significant improvement in the objective was obtained beyond that point. The DSG-based approach always yielded the smallest norm, while the DSG-MQP combination, in particular, generated dual solutions having a significantly smaller dual norm in all test cases with accompanying relatively better or competitive RMP formulations. As far as computational expense is concerned, the proposed DSG-based methods outperformed the CPLEX CG schemes by a factor of 20 for the more difficult data sets (all except PIL67). Note that while Table 2 implies an average speed-up factor of about 10 when using 2,000 DSG iterations in a stand-alone implementation, an embedded DSG scheme takes advantage of an advanced-start capability from the previous iteration, thereby further improving the relative computational performance over several iterations.

The improvement in the final solution quality attained can be attributed to the ability of the DSG scheme to limit dual noise, thereby generating relatively better columns at every CG iteration. This observation is supported by the strong correlation observed in virtually all instances between small dual norms as reported in Column 6 of Table 5 and the LP objective value realized as reported in Column 5.

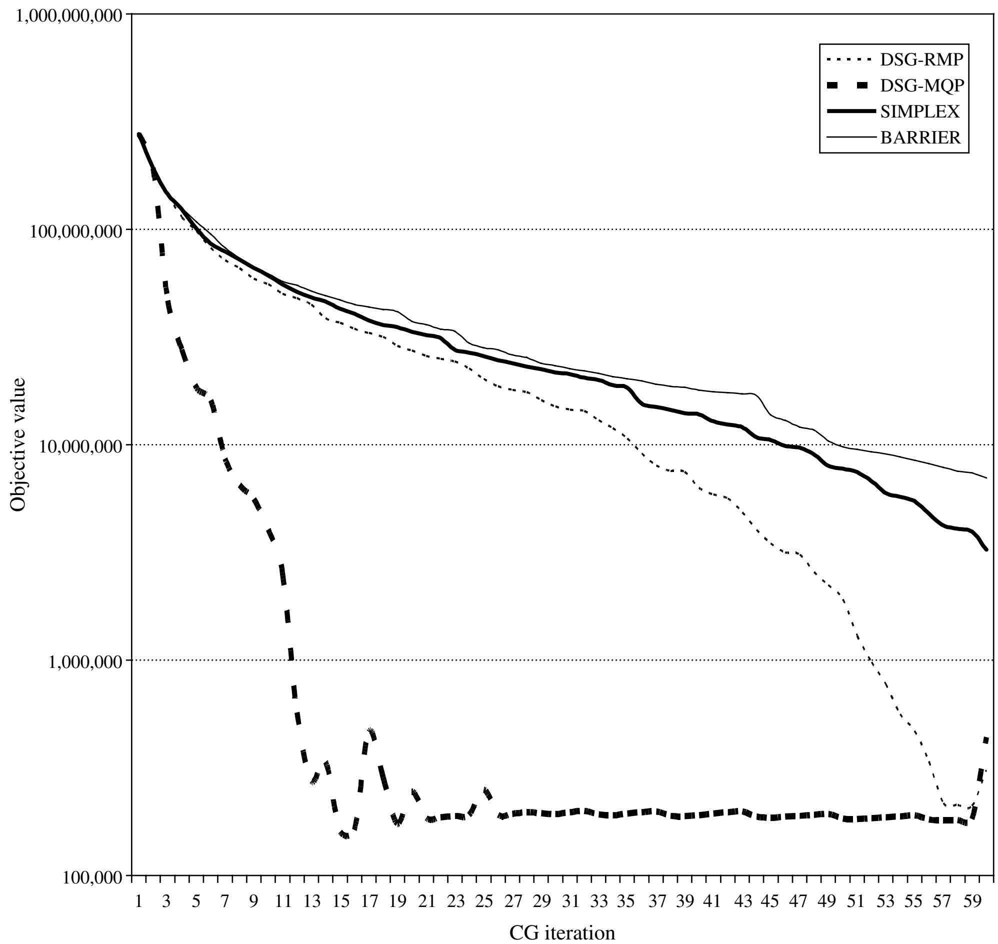


**Figure 2** **Progress in the Objective Value Within ACRUZER Across the CG Iterations for the STALL Data Set**

Furthermore, when we used MQP to seek dual solutions having small norms, we observed a significant improvement in solution quality in almost all examples. In addition, the DSG-RMP solver required about 30% lesser computational time on the average, due to the relatively faster convergence of the RMP objective. In practice, to maintain a good balance between computational performance and solution quality, we recommend using MQP within the CG procedure and to gradually reduce the value of $\omega$, so that we approach a master program formulation that more closely resembles RMP toward the end of the procedure.

Figure 2 compares the trend in the objective value across the CG iterations using each of the aforementioned four approaches for the STALL data set (we display the primal objective value for the CPLEX-based approaches, and $\bar{\theta}$ for the DSG-based approaches except that at the final iteration of the DSG procedures, we record the exact LP value obtained via CPLEX), while Figure 3 compares the value of the dual norm at the corresponding CG iterations. The objective values in Figure 2 are plotted on a logarithmic scale to highlight the substantial performance difference between the CPLEX- and DSG-based approaches. While the DSG-RMP scheme always converged to within the preset tolerance, the DSG-MQP approach terminated with relatively large optimality gaps in four of the 60 CG iterations due to the relatively large value of $\omega$ used in these tests. Consequently, we observe a pronounced nonmonotonic trend in $\bar{\theta}$ and the dual norm for a few iterations in Figures 2 and 3, respectively. However, as noted above, the objective value plotted after the final iteration of the DSG-based schemes is the optimal LP objective value obtained by reoptimizing RMP using CPLEX, resulting in the final "jump" in Figure 2. (The LP values reported in Table 5 are these exact final optimal values.) The trend in these objective values indicates that the CPLEX-based methods exhibit a relatively slow rate of improvement even after 60 CG iterations. The DSG-RMP method converged after 57 iterations, while the DSG-MQP method stabilized after 25 iterations. Note that in the DSG-RMP case, the relatively rapid improvement in the Lagrangian objective after about the 50th iteration in Figure 2 coincides with an equally rapid reduction in dual norm values in Figure 3. Furthermore, the CPLEX-based

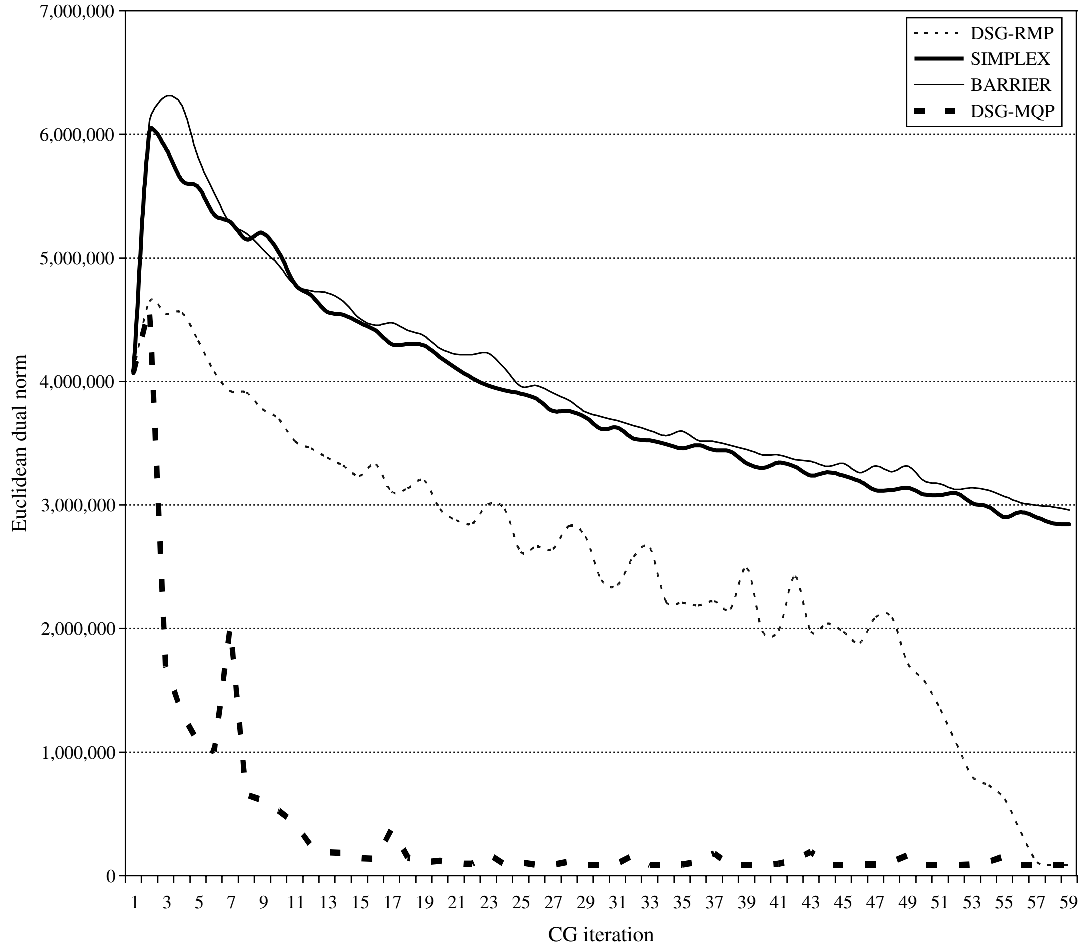


**Figure 3** Variations in the Euclidean Norm of the Dual Solution Across the CG Iterations Within ACRUZER for the STALL Data Set

approaches are characterized by relatively high dual noise throughout the procedure in Figure 3, which correlates with the slow improvement in the objective function value and the relatively large gap from optimality at termination as evident in Figure 2. We observed a similar correlation in the general trends of the objective values and the corresponding dual norms in all the other examples as well.

Finally, we remark on the "production" robustness and versatility of the DSG scheme presented in this paper during industrial use. Proprietary versions of the DSG scheme have been successfully employed within the ACRUZER suite over the last four years and have yielded significant crew cost savings (estimated at \$9.6 million per year and recording the lowest FTC value in history at United Airlines after its initial implementation), accompanied by an equally significant improvement in quality-of-work-life metrics, as well as a reduction in the crew planning cycle time at United Airlines. These successes have resulted in the proposed DSG scheme being additionally deployed without any significant modification within CG-based large-scale aircraft schedule planning applications at United Airlines as well, and has yielded equally successful results.

## 5. Conclusions and Future Research

We have proposed in this paper a new DSG scheme to efficiently generate good solutions to LP relaxations that arise in the context of a CG procedure for solving large-scale crew scheduling problems in practice. In particular, we have identified a phenomenon called *dual noise* that causes a standard commercial-solver-based CG approach implemented at United Airlines to stall far from optimality, and we have demonstrated that the proposed DSG scheme is able to considerably mitigate the effects of the dual noise phenomenon when used to solve the intermediate linear programs that arise within the CG scheme. The DSG scheme possesses several attractive features in comparison with other subgradient methods such as the use of a self-correcting target value, a minimal number of algorithmic parameters that need to be tuned, and empirically observed accelerated convergence to good-quality dual solutions. We have presented sample computational results to demonstrate that the proposed methodology provides a viable alternative to a commercial LP solver within such airline crew planning applications.

Future research includes the possibility of modifying the proposed DSG scheme to derive a stand-alone theoretically convergent method without relying on a safeguarded phase II implementation of a known convergent procedure without compromising its observed effective empirical performance. Although the focus of the present work has been on large-scale crew planning problems, the DSG scheme is equally applicable to solving Lagrangian duals of general linear programs, such as relaxations that arise from 0–1 combinatorial problems in particular. It may also be beneficial to adopt alternative deflection schemes of the type described in Lim and Sherali (2006) and to evaluate the relative performances of these methods on a variety of optimization problems. Our preliminary steps in this direction based on results from test sets from Beasley (1990) have proven to be promising.

## Acknowledgments

This research was supported in part by the National Science Foundation under Grants 0094462 and 0245643. The authors thank Raj Sivakumar, Krishnan Saranathan, and Steve Morley at Enterprise Optimization, United Airlines for their support during this research work, and Pradeep Kumar at Skytech Solutions for his assistance in preparing the test data.

## References

Amor, H. B., J. Desrosiers, J. M. V. Carvalho. 2006. Dual-optimal inequalities for stabilized column generation. *Oper. Res.* **54**(3) 454–463.

Baker, B. M., J. Sheasby. 1999. Accelerating the convergence of subgradient optimisation. *Eur. J. Oper. Res.* **117** 134–144.

Barahona, F., R. Anbil. 2000. The volume algorithm: Producing primal solutions with a subgradient method. *Math. Programming* **87**(3) 385–399.

Barahona, F., R. Anbil. 2002. On some difficult linear programs coming from set partitioning. *Discrete Appl. Math.* **118** 3–11.

Barnes, E., V. Chen, B. Gopalakrishnan, E. L. Johnson. 2002. A least-squares primal-dual algorithm for solving linear programming problems. *Oper. Res. Lett.* **30** 289–294.

Barnhart, C., E. L. Johnson, G. L. Nemhauser, M. W. P. Savelsbergh, P. H. Vance. 1998. Branch-and-price: Column generation for solving huge integer programs. *Oper. Res.* **46**(3) 316–329.

Bazaraa, M. S., H. D. Sherali, C. M. Shetty. 2006. *Nonlinear Programming: Theory and Algorithms*, 3rd ed. John Wiley & Sons, New York.

Beasley, J. 1990. OR-library: Distributing test problems by electronic mail. *J. Oper. Res. Soc.* **41**(11) 1069–1072.

Carvalho, J. M. V. 2005. Using extra dual cuts to accelerate column generation. *INFORMS J. Comput.* **17**(2) 175–182.

Desrosiers, J., Y. Dumas, M. M. Solomon, F. Soumis. 1995. Time constrained routing and scheduling. M. O. Ball, T. L. Magnanti, C. L. Monma, G. L. Nehmhauser, eds. *Handbook in Operations Research/Management Science, Network Routing*, Vol. 8. Elsevier Science, Amsterdam, 35–139.

Elhallaoui, I., D. Villeneuve, F. Soumis, G. Desaulniers. 2005. Dynamic aggregation of set-partitioning constraints in column generation. *Oper. Res.* **53**(4) 632–645.

Gunluk, O., T. Kimbrel, L. Ladanyi, B. Schieber, G. B. Sorkin. 2004. Vehicle routing and staffing for sedan service. http://www.optimization-online.org.

Hjorring, C. 2004. Solving larger crew pairing problems. *Proc. TRISTAN V: The Fifth Triennial Sympos. Transportation Anal., Le Gosier, Guadeloupe*.

Hu, J., E. L. Johnson. 1999. Computational results with a primal-dual subproblem simplex method—A new formulation and decomposition algorithm. *Oper. Res. Lett.* **25** 149–157.

Kelley, J. E. 1960. The cutting-plane method for solving convex programs. *SIAM J. Appl. Math.* **8** 703–712.

Klabjan, D., E. L. Johnson, G. L. Nemhauser, E. Gelman, S. Ramaswamy. 2001. Airline crew scheduling with regularity. *Transportation Sci.* **35**(4) 359–374.

Klabjan, D., E. L. Johnson, G. L. Nemhauser, E. Gelman, S. Ramaswamy. 2002. Airline crew scheduling with time windows and plane-count constraints. *Transportation Sci.* **36**(3) 337–348.

Kohl, N. 2003. Solving the world's largest crew scheduling problem. *ORbit Extra, Danish Oper. Res. Soc.* **4**(Special Issue, August) 8–12.

Lim, C., H. D. Sherali. 2006. Convergence and computational analyses for some variable target value and subgradient deflection methods. *Comput. Optim. Appl.* **34**(3) 409–428.

Makri, A., D. Klabjan. 2004. A new pricing scheme for airline crew scheduling. *INFORMS J. Comput.* **16**(1) 56–67.

Merle, O. D., D. Villeneuve, J. Desrosiers, P. Hansen. 1999. Stabilized column generation. *Discrete Math.* **237** 229–237.

Nemhauser, G. L., L. A. Wolsey. 1999. *Integer and Combinatorial Optimization*, 2nd ed. Wiley Interscience, New York.

Sherali, H. D., G. Choi. 1996. Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs. *Oper. Res. Lett.* **19** 105–113.

Sherali, H. D., C. Lim. 2004. On embedding the volume algorithm in a variable target value method. *Oper. Res. Lett.* **32** 455–462.

Sherali, H. D., G. Choi, C. H. Tuncbilek. 2000. A variable target value method for nondifferentiable optimization. *Oper. Res. Lett.* **26** 1–8.

Vance, P. H., A. Atamurk, C. Barnhart, E. Gelman, E. L. Johnson, A. Krishna, D. Mahidhara, G. L. Nemhauser, R. Rebello. 1997. A heuristic branch-and-price approach for the airline crew pairing problem. Technical Report LEC-97-06, Department of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta.