A decomposition-based warm-start method for stochastic programming*

Marco Colombo † Andreas Grothey‡

School of Mathematics and Maxwell Institute
The University of Edinburgh
Mayfield Road, Edinburgh EH9 3JZ
United Kingdom

Technical Report ERGO 09-008
30 June 2009, revised April 1, 2010

Abstract

In this paper we propose a warm-start technique for interior point methods applicable to multi-stage stochastic programming problems. The main idea is to generate an initial point for the interior point solver by decomposing the barrier problem associated with the deterministic equivalent at the second stage and using a concatenation of the solutions of the subproblems as a starting point for the complete instance. We analyse this scheme and produce theoretical conditions under which the warm-start iterate is successful. We describe the implementation within the OOPS solver and the results of the numerical tests we performed.

1 Introduction

Stochastic Programming Problems [5, 15] are frequently solved by formulating the deterministic equivalent and applying a standard solver to the resulting problem. However for problems formulated over large scenario sets, and in particular for multistage problems, the deterministic equivalent quickly reaches enormous sizes. Due to the special structure of these problems, they can be still be efficiently solved by either approaches based on Benders Decomposition [3, 17, 22] or by structure exploiting Interior Point Methods (IPM) [6, 12, 21]. Common to these approaches is that they obtain the deterministic equivalent by discretising the (in general) continuous distribution of the underlying uncertain parameters. An appealing idea is therefore to work initially with a coarser discretisation of the probability distribution (leading to a smaller problem) and using its solution to crash-start the full problem. Despite seeming an obvious approach, this seems to be a novel idea; probably the reason for this is that both Benders Decomposition and IPM have difficulties in exploiting an advanced crash-start point efficiently. In a previous work

*This research was supported by EPSRC grant EP/E036910/1.
†Email: M.Colombo@ed.ac.uk
‡Email: A.Grothey@ed.ac.uk, Homepage: http://maths.ed.ac.uk/~agr/

[7], the authors have introduced a crash-start approach for stochastic programming based on solving a reduced problem which contains only a trivial fraction of the scenarios. That work provided measurable evidence that a crash-starting scheme based on solving reduced trees within IPMs is workable and leads to measurable efficiency gains.

The approach described in [7] is efficient albeit crude. One of its drawbacks is that the initial point, built from a reduced scenario tree, may be far away from primal-dual feasibility in some of the subproblems. This is particularly the case for trees that are wide and deep, as the representative scenario may not be able to convey all the information of the subtree from which it is chosen. On the other hand, a primal or dual feasible point can be easily constructed by performing a single iteration of a decomposition scheme such as Benders Decomposition or Langragian Relaxation. This has already been used to construct a crash-start point for network flow problems in [10].

In this paper we combine the two ideas: In a first step the problem is solved on a reduced scenario tree to obtain estimates for the first stage decision. These are then used in a second step to solve recourse problems for each scenario, effectively performing half an iteration of Benders Decomposition. From the resulting solutions we construct a crash-start point for the IPM on the full problem. As constructed the resulting crash-start point is always primal feasible. We show that under suitable conditions on the proximity of the reduced scenario tree to the full tree, the IPM warmstart is successful in the sense of [24, 10], that is a full step in the initial iteration can be taken, absorbing any remaining infeasibility. Numerical results show that significant performance gains can be obtained through this scheme.

The paper is structured as follows. In Section 2 we first review relevant background for Stochastic Programming and warm-starting of Interior Point Methods. In Section 3 we introduce the proposed algorithm, which we analyse in Section 4. In Section 5, we present the numerical results obtained with our implementation. Finally, in Section 6 we draw our conclusions and highlight directions of future research.

2 Stochastic Programming and Interior Point Methods

Stochastic programming [5, 15] models uncertainty through the analysis of possible future scenarios. In stochastic programming, the uncertain environment is described through a stochastic process which is obtained from historical data or conjectured according to some prescribed properties. The continuous process is usually further approximated by a discrete distribution in order to obtain a computationally amenable description. This is done by generating a finite, but usually very large, number of scenarios that represent an approximate description of the possible outcomes. The discrete stochastic process can be represented as an *event tree*: each node of the tree denotes a point in time when a realisation of the random process becomes known and a *recourse* decision is taken.

A linear two-stage stochastic programming problem can be formalised as

$$\begin{array}{lll} \min_x & c^\top x + I\!E_\xi[Q(x,\xi)] & \text{where} \quad Q(x,\xi) = \min_y \quad c(\xi)^\top y(\xi) \\ \text{s.t.} & W_0 x = h_0 & \text{s.t.} \quad T x + W(\xi) y(\xi) = h(\xi) \\ & x \geq 0 & \quad\quad y(\xi) \geq 0, \end{array} \tag{1}$$

where the random variable ξ captures the uncertainty. The optimal solution of problem (1)

describes a continuous stochastic process which cannot be solved for directly: as mentioned above, it has to be discretised and rewritten in a form which is viable for computation. For our purposes, we rely on the *deterministic equivalent formulation*. To formulate the deterministic equivalent we adopt the following notation: Let a scenario (π_i, ξ_i) be given by the data $(c_i, W_i, h_i) = (c(\xi_i), W(\xi_i), h(\xi_i))$. A stochastic programming problem is defined by a scenario set (or a *tree*) $\mathcal{T} = \{(\pi_i, \xi_i)_i\}$, so that $P(\mathcal{T})$ denotes the problem

$$\begin{array}{lrll} \min c^\top x + \sum_{i\in\mathcal{T}} \pi_i c_i^\top y_i, \quad \text{s.t.} & W_0 x & = h_0 & \\ & T x + W_i y_i & = h_i, \qquad i \in \mathcal{T} & \qquad (P(\mathcal{T})) \\ & x, y_i & \geq 0. & \end{array}$$

We can introduce a multi-stage decision structure by considering several stages of recourse decisions $(x, y^1, \ldots, y^T)$ and a tree of scenarios. For the purposes of this paper we will assume that the stages $2, \ldots, T$ are represented as one, that is y_i in $(P(\mathcal{T}))$ denotes the vector $y_i = (y_i^1, \ldots, y_i^T)$ and the recourse matrices W_i are structured matrices representing the final $T-1$ stages of the problem.

Several solution methods for stochastic linear programs have been presented in the literature [3, 17, 22]. These often rely on a variant of Benders' decomposition, such as the L-shaped method. They do not require the explicit generation of the deterministic equivalent problem. An entirely different approach, based on interior point methods, exploits the fact that the augmented system matrix arising from applying an IPM to the deterministic equivalent of a large-scale multi-stage stochastic program displays a nested block structure that can be efficiently exploited in the linear algebra.

In this paper we adopt a decomposition-like scheme to build a warm-start iterate that we then use to solve the deterministic equivalent through an interior point method.

2.1 Interior Point Methods

Consider the linear programming problem in standard form

$$\min\ c^\top x \qquad \text{s.t. } Ax = b,\ x \geq 0, \tag{2}$$

where $A \in \mathcal{R}^{m\times n}$ is full rank, $x, c \in \mathcal{R}^n$ and $b \in \mathcal{R}^m$. For the purposes of this paper, problem (2) corresponds to the deterministic equivalent generated from a given event tree $\mathcal{T}$, and we will refer to it as the *complete problem*. Interior point methods work with the barrier problem

$$\min\ c^\top x - \mu \sum \ln x_i \qquad \text{s.t. } Ax = b, \tag{3}$$

instead of (2). Problem (3) is a family of strictly convex problems, parameterised by $\mu > 0$, whose unique solution approaches the solution to the original problem as $\mu \to 0$. The trajectory of solutions to (3) for different values of μ is the *central path*. Interior point methods apply a damped Newton method to the optimality conditions of (3), yielding the direction-finding problem

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^\top & I \\ Z & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} b - Ax \\ c - A^\top y - z \\ -XZe + \mu e \end{bmatrix} = \begin{bmatrix} \xi_b \\ \xi_c \\ \xi_\mu \end{bmatrix}, \tag{4}$$

which needs to be solved with a specified μ for a search direction $(\Delta x, \Delta y, \Delta z)$ at every iteration.

Path-following methods [23] globalise the Newton iteration by keeping the iterates in a neighbourhood of the central path, thus follow it in approaching the optimal solution. A possible choice for the neighbourhood is

$$\mathcal{N}_2(\theta) = \{(x, y, z) : Ax = b, A^T y - z = c, (x, z) > 0, \|XZe - \mu e\|_2 \leq \theta\mu\}, \text{ where } \mu = (x^T z)/n$$

for some $\theta > 0$. For an interior point method to be successful, it is essential that centrality is maintained throughout the iterations, until the optimal partitioning is identified. Approaching a non-optimal vertex too soon can hurt the performance very badly, as the algorithm will strive to recenter the iterate before being able to progress towards optimality. For this reason, the choice of the first iterate is a key issue in an implementation of an interior point method.

In practice, the starting point is generally computed by Mehrotra's starting point heuristic [16], which is considered to be computationally effective. In this heuristic, the starting point is found by solving two least squares problems which attempt to satisfy primal and dual constraints; this point is then shifted away from the boundary towards the positive orthant, in order to satisfy $(x, z) > 0$. Alternatively when a closely related problem has been solved previously, information gained from the solution process of this problem can be used to construct an initial point for the new instance (a process known as warmstarting). Unlike the situation for active set methods (such as the simplex method), when warmstarting IPMs it is not a good idea to use the solution of the previous problem directly as initial iterate. For IPMs this is observed to lead to "jamming" [14, 11], that is the search direction points outside the feasible region leading to very small steps. Theoretical insights [24, 10] suggest that the best warmstart point is near the central path for a not-too-small value of μ. Further a *modification* step is necessary that absorbs any primal-dual infeasibility that may be present in the warmstart point. Let $d = (A, b, c)$ be the data describing the first problem in sequence, and likewise $\bar{d} = (\bar{A}, \bar{b}, \bar{c}) = d + (\Delta A, \Delta b, \Delta c)$ for the second problem that is to be warmstarted. Let $\mathcal{N}_2$ and $\bar{\mathcal{N}}_2$ be corresponding neighbourhoods of the central paths of the two problems. Further we define a norm $\|d\| := \max(\|A\|_2, \|b\|_2, \|c\|_2)$ on the space of problem instances. With $\mathcal{B}$ be the set of ill-posed (infeasible or unbounded) problem instances, and $\rho(d) = \inf\{\|\Delta d\| : d + \Delta d \in \mathcal{B}\}$ the distance to ill-posedness we can define the Renegar condition number

$$C(d) := \|d\|/\rho(d)$$

which has been suggested as a useful indicator of the difficulty of a problem instance when solved by IPM. Then a typical warmstarting result is

Proposition 1 ([24, Prop. 4.2])**.** *Let $(x, y, s) \in \mathcal{N}_2(\theta_0)$ be given with $\mu = x^T s/n$ and suppose that $(\Delta x, \Delta y, \Delta s)$ is obtained by a Weighted-Least-Square modification step. Further let $\theta > \theta_0$, and $\xi \in (0, \theta - \theta_0)$. Assuming that*

$$\delta_{bc} \leq \frac{\theta - \theta_0 - \xi}{(2n+1)\mathcal{C}(d)} \quad \text{and} \quad \mu \geq \frac{\|d\|}{\xi} 4\mathcal{C}(d)^2 \delta_{bc},$$

then $(x + \Delta x, y + \Delta y, s + \Delta s) \in \bar{\mathcal{N}}_2(\theta)$.

Where $\delta_{bc} = (\|\Delta c\|_2 + 2C(d)\|\Delta b\|_2)/\|d\|$. The above result assumes that $\Delta A = 0$, although similar results can be obtained in the general case. A common strategy in practice is to take a point on (or near) the central path for the original problem and use this (after a modification step) as the warmstart point in the new problem. Different modification steps have been suggested in [24, 10]. The theoretical analysis in Section 4 is based on using the Weighted Least Squares (WLS) modification step of [24], although similar results could be obtained for other choices.

3 Decomposition scheme

In the warm-start approach of [7], we build a reduced tree $\mathcal{T}^R$, from the event tree $\mathcal{T}$ associated with the problem, by picking a small number of available scenarios. An approximate solution to the deterministic equivalent corresponding to $\mathcal{T}^R$ is computed and extended to construct a crash-start iterate for the complete problem. The advantage of the scheme, apart from its simplicity, is that the reduced problem is much smaller than the complete formulation, and hence much easier to solve. Still, it provides sufficient information to generate an advanced starting point for the complete problem.

Despite its rather crude nature the scheme of [7] can be surprisingly efficient, even when using a very low number of scenarios in the reduced tree. The reason for this seems to be that for many stochastic programming problems the set of active constraints is very similar for most scenarios. Figure 1 visualises this for some standard two-stage test problems. In these plots each horizontal line corresponds to a scenario: for each variable within a scenario, a pale dot was printed if the variable was active at its lower bound at the optimal solution, a black dot otherwise. Thus, looking vertically, we can spot the difference in the active set between scenarios. It is striking to

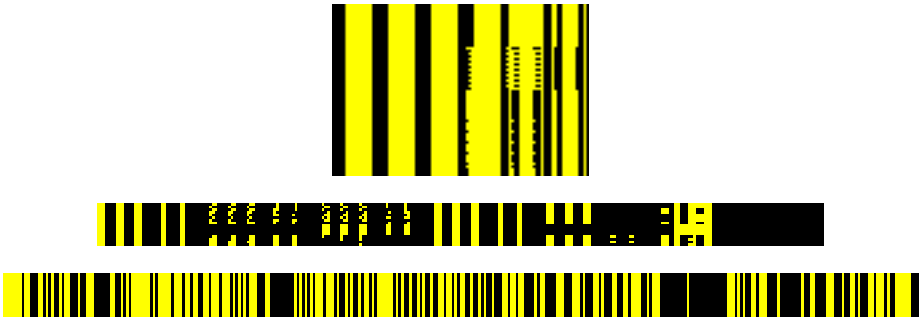

Figure 1: Active set for problem `stocfor2`, `pltexpA2-16` and `fmx2-16`.

see the great similarity among all of them for the problems considered. Many of the scenarios lead to the same active set, and where there are differences these are limited to a small fraction of variables. While the crash-start strategy of [7] is thus typically able to find a good starting point for the majority of scenarios, a bad guess for even a small number of scenarios can severely hamper the performance of the scheme. If a crash-start point is close to a constraint that is not active at the solution, the IPM will need many iterations to move away from such a misidentified active constraint, an effect that has been described as "jamming" or "blocking" [14, 11]. In order to avoid this we will investigate a more sophisticated crash-start procedure.

The idea is based on the observation that if we were given the optimal first stage decision x for problem $(P(\mathcal{T}))$, we could obtain the second stage components y_i by solving a separate sub-problem for each scenario. The underlying idea recalls that of Benders' decomposition.

3.1 A Decomposition-Based Warmstart Scheme

Given a two-stage stochastic programming problem $P(\mathcal{T})$ with the associated scenario set $\mathcal{T}$, we start by considering the corresponding barrier problem:

$$\begin{array}{lll} \min & c^\top x + \sum_{i\in\mathcal{T}} \pi_i c_i^\top y_i - \mu \sum_j \ln x_j - \mu \sum_i \sum_j \ln y_{i,j} & \\ \text{s.t.} & Tx + W_i y_i = h_i, \qquad i \in \mathcal{T} & (P_\mu(\mathcal{T})) \\ & x, y_i \geq 0. & \end{array}$$

and its KKT conditions

$$s + \sum_i T^\top \lambda_i = c \tag{5a}$$
$$z_i + W_i^\top \lambda_i = \pi_i c_i \tag{5b}$$
$$Tx + W_i y_i = h_i \tag{5c}$$
$$XSe = \mu e \tag{5d}$$
$$Y_i Z_i e = \mu e \tag{5e}$$

For a given value $\hat{x}$ of the first stage decisions, the barrier problem $P_\mu(\mathcal{T})$ decomposes into the scenario subproblems

$$Q_{i,\mu}(\hat{x}) = \min\{\pi_i c_i^\top y_i - \mu \sum_j \ln y_{i,j} : W_i y_i = h_i - T_i \hat{x},\ y_i \geq 0\}, \qquad (P_{i,\mu}(\hat{x}))$$

with optimality conditions

$$W_i^\top \lambda_i + z_i = \pi_i c_i \tag{6a}$$
$$W_i y_i = h_i - T\hat{x} \tag{6b}$$
$$Y_i Z_i e = \mu e \tag{6c}$$

which together cover the conditions (5b),(5c) and (5e). Conditions (5a) and (5d) on the other hand reflect on the centrality and optimality of the chosen $\hat{x}$.

Thus, if we are given a reasonable central and optimal $\hat{x}, \hat{s}$ that are good approximations to the x, s-components of the solution to $P_\mu(\mathcal{T})$, we can use this to solve subproblems $Q_{i,\mu}(\hat{x})$. Combining $\hat{x}, \hat{s}$ with the thus obtained $(y_i^*, z_i^*, \lambda_i^*)$ for all subproblems yields an primal feasible (5a) and central (5d,e) approximation $(\bar{x}, \bar{s}, \bar{y}, \bar{z}, \bar{\lambda}) = (\hat{x}, \hat{s}, y^*, z^*, \lambda^*)$ to the solution of $P_\mu(\mathcal{T})$. Dual feasibility will depend on the quality of the chosen $(\hat{x}, \hat{s})$. In the suggested scheme we will obtain estimates $(\hat{x}, \hat{s})$ by solving $P_\mu(\mathcal{T}^R)$ for a reduced tree $\mathcal{T}^R$. We will show that for if $\mathcal{T}^R$ approximates $\mathcal{T}$ sufficienty well, the constructed warmstart point $(\bar{x}, \bar{s}, \bar{y}, \bar{z}, \bar{\lambda})$ will be in an appropriate N_2-neighbourhood for problem $P_\mu(\mathcal{T})$ and thus a sufficiently good warmstart point by Proposition 1. The resulting decomposition based crash-start algorithm is summarized as Algorithm 1.

Compared to the approach of [7], this scheme provides directly the part of the warm-start vector for every node in the tree, and so it does not require a process of "expanding the solution" by copying parts of solution from extraneous nodes. As such, we expect it to provide an iterate that is closer to the central path of the complete problem: by construction the warm-start point is central and primal feasible, while dual feasibility depends on the quality of the estimate of the first-stage decisions. We will back up these claims in the following section.

Algorithm 1 Decomposition-based warm-start algorithm

Require: The complete event tree $\mathcal{T}$.

1: Generate a reduced event tree $\mathcal{T}^R \subset \mathcal{T}$; Fix a target $\bar{\mu} > 0$.
2: Solve problem the barrier problem on the reduced tree $P_{\bar{\mu}}(\mathcal{T}^R)$ and obtain first stage decisions and duals $(\hat{x}, \hat{s})$ from it.
3: **for** $i \in \{$second stage nodes of $\mathcal{T}\}$ **do**
4: Solve scenario subproblem $Q_{i,\bar{\mu}}(\bar{x})$ to obtain recourse decisions and duals $(y_i^*, z_i^*, \lambda_i^*)$.
5: **end for**
6: Combine these to a crash start iterate $(\bar{x}, \bar{s}, \bar{y}, \bar{z}, \bar{\lambda})$ for the complete problem $P(\mathcal{T})$.
7: Perform a modification step in the full problem and solve to optimimality.

4 Theoretical analysis

The aim of this section is to state conditions under which the decomposition-based crash-start procedure is successful. That is we are able to construct a point that is sufficiently primal–dual feasible and central to satisfy the conditions of Proposition 1. We will assume that the problem is a two-stage problem.

We assume that we have full recourse, so that every scenario subproblem is feasible and bounded for all x. Indeed for the technical results we need to assume more: following the terminology of Section 2.1 we use $d_i(\hat{x}) = (W_i, h_i - T\hat{x}, c_i)$ to denote the problem data of the i-th scenario subproblem and define the distance to ill-posedness for problem $P_i(x)$ as

$$\rho(d_i(x)) := \inf\{\|\Delta d\| : d_i + \Delta d \in \mathcal{B}\},$$

with $\mathcal{B}$ being the set of "ill-posed" data instances, and we require that

$$\underline{\rho} := \inf_x \min_{i \in \mathcal{T}} \rho(d_i(x)) > 0.$$

Further we assume that the problem data itself is bounded, i.e. $\|d_i\| \leq \overline{\|d\|}$, which gives us the existence of a bound $\overline{C(d)}$ on the Nunez-Freund condition number $C(d) = \|d\|/\rho(d)$:

$$C(d_i) \leq \overline{C(d)}, \quad \forall i \in \mathcal{T}.$$

In what follows we deal with three types of problems: The full problem $P(\mathcal{T})$, the reduced problem $P(\mathcal{T}^R)$ and the scenario subproblems $P_i(\hat{x})$. We use the convention that objects associated with the full problem, the reduced problem and the scenario subproblems have superscripts ${}^{\mathcal{T}}$, ^{R} and ^{P_i} respectively (as in $\mathcal{N}_2^{\mathcal{T}}, \mathcal{N}_2^R, \mathcal{N}_2^{P_i}$). We will also use the notation such as $x_\mu(\mathcal{T})$ to refer to the x-component of the solution to problem $P_\mu(\mathcal{T})$.

In our decomposition-based warm-start method, starting from the problem of interest $P(\mathcal{T})$, we identify a reduced tree $\mathcal{T}^R$, and find a $\bar{\mu}$-central point $(x^R, y^R, \lambda^R, s^R, z^R) \in \mathcal{N}_2^R(\theta_0)$. We take the x, s-components of this point as our estimate of the optimal first-stage decisions to solve the subproblems $P_i(x), \forall i \in \mathcal{T}$, obtaining $\bar{\mu}$-central points

$$(y_i^*, \lambda_i^*, z_i^*) \in \mathcal{N}_2^{P_i}(\theta_0), \quad \forall i \in \mathcal{T}.$$

Then (x^R, s^R) and $(y_i^*, \lambda_i^*, z_i^*)_{i \in \mathcal{T}}$ are combined to obtain the *warm-start point* $\bar{w} = (\bar{x}, \bar{s}, \bar{y}, \bar{z}, \bar{\lambda})$. We aim to show that the warmstart from this point in problem $P(\mathcal{T})$ is successful, in that the modification step $\Delta w = (\Delta x, \Delta s, \Delta y, \Delta z, \Delta\lambda)$ from $\bar{w}$ satisfies

$$\bar{w} + \Delta w \in \mathcal{N}_2^{\mathcal{T}}(\theta).$$

In particular we will show that the warm-start is successful if the reduced tree $\mathcal{T}^R$ is close enough to $\mathcal{T}$ in an appropriate measure.

As the measure of closeness of the trees we use the *Wasserstein distance* [9, 19]. In an abstract probability space setting with event space Ω and corresponding Borel σ-field $\mathcal{B}$, the Wasserstein distance of two probability measures $\nu, \tilde{\nu}$ on $(\Omega, \mathcal{B})$ is given by

$$\hat{\mu}_1(\nu, \tilde{\nu}) = \inf \left\{ \int_{\Omega \times \Omega} \|\omega - \tilde{\omega}\| \eta(\mathrm{d}(\omega, \tilde{\omega})) : \begin{array}{l} \eta \in \mathcal{P}(\Omega \times \Omega), \eta(B \times \Omega) = \nu(B), \\ \eta(\Omega \times B) = \tilde{\nu}(B), \forall B \in \mathcal{B} \end{array} \right\}.$$

In our case, where ν and $\tilde{\nu}$ are discrete measures implied by the full and reduced scenario trees as

$$\nu = \sum_{(\pi_i, \xi_i) \in \mathcal{T}} \pi_i \delta_{\xi_i}, \qquad \tilde{\nu} = \sum_{(\tilde{\pi}_i, \tilde{\xi}_i) \in \mathcal{T}^{\mathcal{R}}} \tilde{\pi}_i \delta_{\tilde{\xi}_i},$$

where δ_{ξ_i} is the Dirac-measure placed on scenario ξ_i, the corresponding formula for the Wasserstein distance reduces to the transportation distance

$$W_1(\mathcal{T}, \mathcal{T}^R) = \hat{\mu}_1(\nu, \tilde{\nu}) = \min_{\eta \geq 0} \left\{ \sum_{(\pi_i, \xi_i) \in \mathcal{T}} \sum_{(\tilde{\pi}_j, \tilde{\xi}_j) \in \mathcal{T}^R} \|\xi_i - \tilde{\xi}_j\| \eta_{ij} : \sum_i \eta_{ij} = \tilde{\pi}_j, \sum_j \eta_{ij} = \pi_i \right\}.$$

Our analysis first deals with an idealised algorithm in which we choose a target μ value of $\bar{\mu}$ and then obtain the exact μ-center for reduced problem $P(\mathcal{T}^R)$ and scenario subproblems $P_i(x)$. Later we give results for the more realistic algorithm in which we are content with finding points in a neighbourhood of the central path. In what follows we assume that the random parameter ξ only affects the right-hand side h of the problem, and that this influence is bounded by the Lipschitz constant H:

$$\|h(\xi) - h(\tilde{\xi})\| \leq H \|\xi - \tilde{\xi}\|, \qquad \forall \xi, \tilde{\xi}.$$

Moreover, we make use of the following relation from Dikin [8]

$$\chi(W) := \sup_{\Sigma \in D^+} \|\Sigma W^\top (W \Sigma W^\top)^{-1}\|_\infty < \infty, \tag{7}$$

where D^+ is the set of all diagonal matrices with strictly positive diagonal elements and W is an arbitrary matrix.

4.1 Results for exact subproblem solutions

We first investigate the dependence of the value functions of the family of barrier problems (for different values of ν)

$$\begin{array}{ll} \hat{Q}_\nu(x; \xi) = \min & c^\top y - \nu \sum_j \ln y_j \\ & \text{s.t. } Wy = h(\xi) - Tx \end{array} \tag{8}$$

on the random parameter ξ. We start with the following Lemma with establishes a Lipschitz result for the function $Q_\nu(x; \xi)$.

Lemma 1. *The functions $\hat{Q}_\nu(x; \xi)$ are Lipschitz in the second argument, that is*

$$|\hat{Q}_\nu(x; \xi) - \hat{Q}_\nu(x; \bar{\xi})| \leq L_Q \|\xi - \bar{\xi}\|_2$$

with Lipschitz constant $L_Q := 3\overline{C(d)}(\overline{C(d)\|d\|} + \nu n)\chi(W)H$.

Proof. Optimality conditions for (8) are

$$z + W^\top \lambda = c, \qquad Wy + Tx = h(\xi), \qquad YZe = \nu e. \tag{9}$$

Under our assumptions, this system is non-singular, so the implicit function theorem assures the existence of y, λ and s as differentiable functions of ξ. Differentiating with respect to ξ gives

$$W\frac{\mathrm{d}y}{\mathrm{d}\xi} = \frac{\mathrm{d}h}{\mathrm{d}\xi} \tag{10a}$$

$$\frac{\mathrm{d}z}{\mathrm{d}\xi} + W^\top \frac{\mathrm{d}\lambda}{\mathrm{d}\xi} = 0 \tag{10b}$$

$$Y\frac{\mathrm{d}z}{\mathrm{d}\xi} + Z\frac{\mathrm{d}y}{\mathrm{d}\xi} = 0. \tag{10c}$$

After rearranging the final equation for $\frac{\mathrm{d}z}{\mathrm{d}\xi}$ and substituting into the second equation we obtain

$$-Y^{-1}Z\frac{\mathrm{d}y}{\mathrm{d}\xi} + W^\top \frac{\mathrm{d}\lambda}{\mathrm{d}\xi} = 0,$$

and hence

$$\frac{\mathrm{d}y}{\mathrm{d}\xi} = YZ^{-1}W^\top \frac{\mathrm{d}\lambda}{\mathrm{d}\xi}. \tag{11}$$

Multiplying from the left with W and substituting into (10a) yields

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\xi} = (WYZ^{-1}W^\top)^{-1}\frac{\mathrm{d}h}{\mathrm{d}\xi},$$

which together with (11) gives

$$\frac{\mathrm{d}y}{\mathrm{d}\xi} = YZ^{-1}W^\top(WYZ^{-1}W^\top)^{-1}\frac{\mathrm{d}h}{\mathrm{d}\xi}.$$

Now, recalling (7), we get the bound

$$\left\|\frac{\mathrm{d}y}{\mathrm{d}\xi}\right\|_\infty \leq \chi(W)\left\|\frac{\mathrm{d}h}{\mathrm{d}\xi}\right\|_\infty.$$

On the other hand, from the definition of Q_ν in (8) we have

$$\frac{\mathrm{d}Q_\nu}{\mathrm{d}\xi} = c^\top\frac{\mathrm{d}y}{\mathrm{d}\xi} - \nu Y^{-1}\frac{\mathrm{d}y}{\mathrm{d}\xi} = (c-z)^\top\frac{\mathrm{d}y}{\mathrm{d}\xi}.$$

From [18, Theorem 3.1] we have the bound

$$\|z\|_\infty \leq 2C(d)(C(d)\|d\| + \nu n),$$

which together with $\|c\|_\infty \leq \|d\|$ and $\|\frac{\mathrm{d}h}{\mathrm{d}\xi}\|_\infty \leq H$ yields

$$\left\|\frac{\mathrm{d}Q_\nu}{\mathrm{d}\xi}\right\|_\infty \leq \left[\|d\| + 2C(d)(C(d)\|d\| + \nu n)\right]\chi(W)H,$$

and the assertion of the Lemma follows since $C(d) \geq 1$. □

In what follows we define

$$\eta_{\mathcal{T}}(x) := c^\top x - \mu \sum_j \ln x_j + \rho_{\mathcal{T}}(x) \tag{12}$$

where

$$\rho_{\mathcal{T}}(x) = \sum_{(\pi_i,\xi_i)\in\mathcal{T}} \pi_i Q_{\frac{\mu}{\pi_i}}(x;\xi_i).$$

We are interested in how the optimal first-stage solution $x_\mu(\mathcal{T})$ to the problem

$$\min_x \eta_{\mathcal{T}}(x) \tag{$P_\mu(\mathcal{T})$}$$

depends on the underlying scenario set $\mathcal{T}$. We need to assume that the set of possible optimal first-stage decisions x for different trees is bounded, that is

$$\|x_\mu(\mathcal{T})\|_\infty \le \overline{B}, \quad \text{for all considered } \mathcal{T}.$$

Also we assume that the probabilities of all considered scenarios in the full and reduced tree are bounded below by $\bar{\pi}$. As long as we only consider trees that are derived from the original tree by deletion of scenarios and aggregation, both assumptions hold trivially, since there is only a finite number of possible trees to consider.

Lemma 2. *With the notation introduced above we have*

$$\|x_\mu(\mathcal{T}) - x_\mu(\mathcal{T}^R)\|_\infty \le \frac{2}{\sqrt{\mu}}\overline{B}\sqrt{L_Q(\mu/\bar{\pi})}\sqrt{W_1(\mathcal{T},\mathcal{T}^R)}$$

and

$$\|s_\mu(\mathcal{T}) - s_\mu(\mathcal{T}^R)\|_\infty \le \frac{2}{\sqrt{\mu^3}}\overline{B}^3\sqrt{L_Q(\mu/\bar{\pi})}\sqrt{W_1(\mathcal{T},\mathcal{T}^R)}.$$

Proof. We make use of Remark 2.2 to [19, Proposition 2.1] (see also [20, Theorem 2.10]), which states that given $|h(z) - h(\tilde{z})| \le L_h\|z - \tilde{z}\|$, we have

$$\left|\int h(z)\nu(\mathrm{d}z) - \int h(z)\tilde{\nu}(\mathrm{d}z)\right| \le L_h W_1(\nu,\tilde{\nu}).$$

We apply this result to the function $Q_\mu(x;\xi)$, with $\nu,\tilde{\nu}$, as before, being the probability measures implied by the full and reduced trees. In this case

$$\left|\sum_{i\in\mathcal{T}} \pi_i Q_{\frac{\mu}{\pi_i}}(x,\xi_i) - \sum_{i\in\mathcal{T}^R} \tilde{\pi}_i Q_{\frac{\mu}{\tilde{\pi}_i}}(x,\tilde{\xi}_i)\right| \le L_Q(\mu/\bar{\pi})W_1(\mathcal{T},\mathcal{T}^R)$$

therefore

$$|\rho_{\mathcal{T}}(x) - \rho_{\mathcal{T}^R}(x)| \le L_Q W_1(\mathcal{T},\mathcal{T}^R),$$

and hence also

$$|\eta_{\mathcal{T}}(x) - \eta_{\mathcal{T}^R}(x)| \le L_Q W_1(\mathcal{T},\mathcal{T}^R). \tag{13}$$

By $x_\mu(\mathcal{T}^R)$ being a minimizer of $\eta_{\mathcal{T}^R}(x)$, it holds that

$$\eta_{\mathcal{T}^R}(x_\mu(\mathcal{T}^R)) \le \eta_{\mathcal{T}^R}(x_\mu(\mathcal{T})). \tag{14}$$

Bound (13) implies

$$\eta_{\mathcal{T}^R}(x_\mu(\mathcal{T})) \leq \eta_{\mathcal{T}}(x_\mu(\mathcal{T})) + L_Q W_1(\mathcal{T}, \mathcal{T}^R),$$

and hence together with (14)

$$\eta_{\mathcal{T}^R}(x_\mu(\mathcal{T}^R)) \leq \eta_{\mathcal{T}}(x_\mu(\mathcal{T})) + L_Q W_1(\mathcal{T}, \mathcal{T}^R). \tag{15}$$

The functions $\rho_{\mathcal{T}}(x), \rho_{\mathcal{T}^R}(x)$ are convex and differentiable, and the term $-\mu \sum_j \ln x_j$ is convex and twice continuously differentiable with Hessian

$$\nabla_x^2[-\mu \sum_j \ln x_j] = \mu X^{-2}$$

and its lowest eigenvalue satisfies

$$\sigma_1(\mu X^{-2}) \geq \mu/(\max_j x_j)^2 \geq \mu/\overline{B}^2.$$

Hence we get the bound

$$\eta_{\mathcal{T}}(x_\mu(\mathcal{T}^R)) \geq \eta_{\mathcal{T}}(x_\mu(\mathcal{T})) + \nabla_x \eta_{\mathcal{T}}(x_\mu(\mathcal{T}))^\top (x_\mu(\mathcal{T}^R) - x_\mu(\mathcal{T})) + \frac{1}{2}\|x_\mu(\mathcal{T}^R) - x_\mu(\mathcal{T})\|_\infty^2 \frac{\mu}{\overline{B}^2}.$$

Since $\eta_{\mathcal{T}}(x)$ is convex and differentiable with minimizer $x_\mu(\mathcal{T})$, then $\nabla_x \eta_{\mathcal{T}}(x_\mu(\mathcal{T})) = 0$, so that

$$\eta_{\mathcal{T}}(x_\mu(\mathcal{T}^R)) \geq \eta_{\mathcal{T}}(x_\mu(\mathcal{T})) + \frac{1}{2}\|x_\mu(\mathcal{T}^R) - x_\mu(\mathcal{T})\|_\infty^2 \frac{\mu}{\overline{B}^2}. \tag{16}$$

Therefore, by combining (13) and (16) we get

$$\begin{aligned} \eta_{\mathcal{T}^R}(x_\mu(\mathcal{T}^R)) &\geq \eta_{\mathcal{T}}(x_\mu(\mathcal{T}^R)) - L_Q W_1(\mathcal{T}, \mathcal{T}^R) \\ &\geq \eta_{\mathcal{T}}(x_\mu(\mathcal{T})) + \frac{1}{2}\|x_\mu(\mathcal{T}^R) - x_\mu(\mathcal{T})\|_\infty^2 \frac{\mu}{\overline{B}^2} - L_Q W_1(\mathcal{T}, \mathcal{T}^R). \end{aligned}$$

On the other hand, the minimum value $\eta_{\mathcal{T}^R}(x_\mu(\mathcal{T}^R))$ of $P_\mu(\mathcal{T}^R)$ needs to satisfy (15) so that we have

$$\eta_{\mathcal{T}}(x_\mu(\mathcal{T})) + L_Q W_1(\mathcal{T}, \mathcal{T}^R) \geq \eta_{\mathcal{T}}(x_\mu(\mathcal{T})) + \frac{1}{2}\|x_\mu(\mathcal{T}^R) - x_\mu(\mathcal{T})\|_\infty^2 \frac{\mu}{\overline{B}^2} - L_Q W_1(\mathcal{T}, \mathcal{T}^R).$$

After rearrangement we are left with

$$\frac{1}{2}\|x_\mu(\mathcal{T}^R) - x_\mu(\mathcal{T})\|_\infty^2 \frac{\mu}{\overline{B}^2} \leq 2L_Q W_1(\mathcal{T}, \mathcal{T}^R),$$

or equivalently

$$\|x_\mu(\mathcal{T}) - x_\mu(\mathcal{T}^R)\|_\infty \leq \frac{2}{\sqrt{\mu}} \overline{B} \sqrt{L_Q} \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)}.$$

For the bound on $\|s_\mu(\mathcal{T}) - s_\mu(\mathcal{T}^R)\|_\infty$ we note that

$$X_\mu(\mathcal{T}) S_\mu(\mathcal{T}) = \mu I, \qquad X_\mu(\mathcal{T}^R) S_\mu(\mathcal{T}^R) = \mu I,$$

so that

$$s_\mu(\mathcal{T}) = \frac{1}{\mu} S_\mu(\mathcal{T}) S_\mu(\mathcal{T}^R) x_\mu(\mathcal{T}^R), \quad s_\mu(\mathcal{T}^R) = \frac{1}{\mu} S_\mu(\mathcal{T}) S_\mu(\mathcal{T}^R) x_\mu(\mathcal{T}),$$

and hence

$$\begin{aligned}\|s_\mu(\mathcal{T}) - s_\mu(\mathcal{T}^R)\|_\infty &= \|\frac{1}{\mu} S_\mu(\mathcal{T}) S_\mu(\mathcal{T}^R)(x_\mu(\mathcal{T}^R) - x_\mu(\mathcal{T}))\|_\infty \\ &\leq \frac{\overline{B}^2}{\mu} \|x_\mu(\mathcal{T}) - x_\mu(\mathcal{T}^R))\|_\infty,\end{aligned}$$

yielding the required bound. □

Lemma 3. *Let estimates of the first-stage decisions $\bar{x} = x^R, \bar{s} = s^R$ be given and let (y_i, λ_i, z_i) be the (exact) μ-center for the i-th scenario subproblem $P_i(\bar{x})$. Further, let $(x_\mu, y_\mu, \lambda_u, s_\mu, z_\mu)$ be the exact μ-center of the full problem $P(\mathcal{T})$. Then*

$$\|\lambda_i - \lambda_{\mu,i}\|_\infty \leq C_\lambda \|\bar{x} - x_\mu\|_\infty$$

where

$$C_\lambda = C_\lambda(\mu) = 4\chi(W)\|T\|_\infty \overline{C(d)}^2 [\overline{C(d)}\|d\| + \mu n]^2/\mu.$$

Proof. The optimality conditions for problem $P_i(\bar{x})$ are given by (9). Differentiating with respect to x gives:

$$\frac{\mathrm{d}z_i}{\mathrm{d}x} + W^\top \frac{\mathrm{d}\lambda_i}{\mathrm{d}x} = 0, \qquad W\frac{\mathrm{d}y_i}{\mathrm{d}x} = -T, \qquad Y_i \frac{\mathrm{d}z_i}{\mathrm{d}x_i} + Z_i \frac{\mathrm{d}y_i}{\mathrm{d}x} = 0.$$

As in the proof to Lemma 1, these equations can be solved for $\frac{\mathrm{d}\lambda_i}{\mathrm{d}x}$ to obtain

$$\begin{aligned}\frac{\mathrm{d}\lambda_i}{\mathrm{d}x} &= (WY_iZ_i^{-1}W^\top)^{-1}T \\ &= (WW^\top)^{-1}WY_i^{-1}Z_iY_iZ_i^{-1}W^\top(WY_iZ_i^{-1}W^\top)^{-1}T.\end{aligned}$$

As before we have

$$\|Z_i^{-1}W^\top(WY_iZ_i^{-1}W^\top)^{-1}\|_\infty \leq \chi(W), \qquad \|W^\top(WW^\top)^{-1}\|_\infty \leq \chi(W),$$

and, due to (y_i, λ_i, z_i) being the exact μ-center,

$$\|Z_iY_i^{-1}\|_\infty = \|Z_i^2/\mu\|_\infty = \|Z_i\|_\infty^2/\mu.$$

We use again the bound from [18, Theorem 3.1]

$$\|z_i\|_\infty \leq 2C(d_i)[C(d_i)\|d_i\| + \mu n]$$

giving

$$\begin{aligned}\left\|\frac{\mathrm{d}\lambda_i}{\mathrm{d}x}\right\|_\infty &\leq 4\chi(W)^2\|T\|_\infty C(d_i)^2[C(d_i)\|d_i\| + \mu n]^2/\mu \\ &\leq 4\chi(W)^2\|T\|_\infty \overline{C(d)}^2[\overline{C(d)\|d\|} + \mu n]^2/\mu\end{aligned}$$

yielding the assertion of the Lemma. □

Theorem 4. *Let $\bar{w} = (\bar{x}, \bar{y}, \bar{\lambda}, \bar{s}, \bar{z})$ be the warm-start point for problem $P(\mathcal{T})$ obtained by following the above algorithm starting from the reduced tree $\mathcal{T}^R$, using a target μ-value of $\bar{\mu}$. If*

$$W_1(\mathcal{T}, \mathcal{T}^R) \leq \frac{\theta^2}{\overline{C(d)}^2 C_\lambda(\mu)^2} \min\left\{ \frac{\overline{\|d\|}^2}{4(2n+1)^2}, \frac{\mu^2}{16\overline{C(d)}^2} \right\}$$

then the warmstart is successful, that is the Yıldırım -Wright [24] Weighted Least Squares step $(\Delta x, \Delta y, \Delta\lambda, \Delta s)$ from $(\bar{x}, \bar{y}, \bar{\lambda}, \bar{s})$ is feasible and leads to

$$(\bar{x} + \Delta x, \bar{y} + \Delta y, \bar{\lambda} + \Delta\lambda, \bar{s} + \Delta s) \in \mathcal{N}_2^{\mathcal{T}}(\theta).$$

Proof. Due to the construction of the warm-start point $\bar{w}$ we have

$$T\bar{x} + W\bar{y}_i = h_i, \qquad \bar{X}\bar{S}e = \mu e, \qquad \bar{Y}_i\bar{Z}_i e = \mu e,$$

that is, the point is primal feasible and central in the complete problem $P(\mathcal{T})$. Let

$$\bar{c} = \sum_i T^\top \bar{\lambda}_i + \bar{s}$$

and consider the problem instance $P(\bar{d})$ obtained from the full problem $P(\mathcal{T})$ by replacing the first-stage cost c with $\bar{c}$. By construction, $\bar{w}$ is also dual feasible for $P(\bar{d})$, hence it is the exact μ-center, or $\bar{w} \in \bar{\mathcal{N}}_2(0)$, where $\bar{\mathcal{N}}_2$ is the $\mathcal{N}_2$-neighbourhood for problem $P(\bar{d})$. We will treat this point as a warm-start attempt for the (perturbed) problem $P(\mathcal{T})$ starting from a central point for problem $P(\bar{d})$. The change in problem data is

$$\Delta d = (\Delta A, \Delta b, \Delta c) = (0, 0, \Delta c),$$

with

$$\begin{aligned} \Delta c = c - \bar{c} &= \sum_i T^\top \lambda_{\mu,i} + s_\mu - \sum_i T^\top \bar{\lambda}_i - \bar{s} \\ &= \sum_i T^\top (\lambda_{\mu,i} - \bar{\lambda}_i) + (s_\mu - \bar{s}). \end{aligned}$$

Using the bounds from Lemma 2 and Lemma 3 we have

$$\begin{aligned} \|\Delta c\|_\infty &\leq \|T\|_\infty |\mathcal{T}| \max_i \|\lambda_{\mu,i} - \bar{\lambda}_i\|_\infty + \|s_\mu - \bar{s}\|_\infty \\ &\leq \|T\|_\infty |\mathcal{T}| C_\lambda(\mu) \|x_\mu - \bar{x}\|_\infty + \|s_\mu - \bar{s}\|_\infty \\ &\leq \left(\|T\|_\infty |\mathcal{T}| C_\lambda(\mu) + \frac{\overline{B}^2}{\mu} \right) \frac{2}{\sqrt{\mu}} \overline{B} \sqrt{L_Q} \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} \\ &= C_2(\mu) \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} \end{aligned}$$

with

$$C_2(\mu) = \left(\|T\|_\infty |\mathcal{T}| C_\lambda(\mu) + \frac{\overline{B}^2}{\mu} \right) \frac{2}{\sqrt{\mu}} \overline{B} \sqrt{L_Q}.$$

According to [24, Proposition 4.2] sufficient conditions for a successful warmstart are (using $\theta_0 = 0$, $\xi = \theta/2$):

$$\|\Delta c\|_\infty \leq \frac{\theta}{2(2n+1)C(d)} \|d\|, \qquad \text{and} \qquad \mu \geq 8 \frac{C(d)^2}{\theta} \|\Delta c\|_\infty,$$

which can be combined to produce

$$\|\Delta c\|_\infty \leq \frac{\theta}{C(d)} \min\left\{\frac{\|d\|}{2(2n+1)}, \frac{\mu}{8C(d)}\right\}.$$

Together with the above bound on $\|\Delta c\|_\infty$, we obtain the condition

$$C_2(\mu)\sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} \leq \frac{\theta}{C(d)} \min\left\{\frac{\|d\|}{2(2n+1)}, \frac{\mu}{8C(d)}\right\},$$

and, after rearranging and substituting in for $C_2(\mu)$, the condition given in the Theorem. □

The question remains how the condition of the Theorem can be satisfied in practice. Clearly the bound on $W_1(\mathcal{T}, \mathcal{T}^R)$ is largest when the two expressions in the min are equal, that is if

$$\frac{\|d\|}{2(2n+1)} = \frac{\mu}{8C(d)} \qquad \text{or} \qquad \mu = \frac{4\|d\|C(d)}{2n+1}.$$

For an (optimally) chosen μ-value, Theorem 4 gives conditions on the closeness of the trees $\mathcal{T}$ and $\mathcal{T}^R$. It is not practically possible to determine this value exactly, so the μ value would in practice be chosen by an appropriate heuristic.

Closer to the practical application is probably to choose the two trees first and then the corresponding μ value to be used in the solution of the reduced problem $P(\mathcal{T}^R)$ and the subproblems $P_i(\bar{x})$. In this view of things, Theorem 4 gives a minimum closeness of the approximating tree $\mathcal{T}^R$ that needs to be achieved and also gives conditions on the selection of the corresponding μ value. In practice we expect the conditions to be much too tight and will proceed with the warm-start attempt regardless.

4.2 Results for approximate subproblem solutions

The requirement to obtain an exact μ-center for the reduced problem $P(\mathcal{T}^R)$ or the decomposed subproblems $P_i(\bar{x})$ is too demanding. Rather we will be satisfied with obtaining primal–dual feasible and reasonably central points, i.e.

$$(\tilde{x}_\mu^R, \tilde{y}_\mu^R, \tilde{\lambda}_\mu^R, \tilde{s}_\mu^R, \tilde{z}_\mu^R) \in \mathcal{N}_2^R(\theta) \tag{17}$$

$$(\tilde{y}_i, \tilde{\lambda}_i, \tilde{z}_i) \in \mathcal{N}_2^{(i)}(\theta) \tag{18}$$

for some $\theta \in (0, 1)$. The following Lemma 6 gives bounds on the resulting error in the relevant components of these points compared to the exact μ-centers. Before we proceed, we need a general result stating how far the components of a point in the $\mathcal{N}_2$-neighbourhood can deviate from the exact μ-center.

Lemma 5. *Let (x_μ, y_μ, s_μ) be the exact μ-center for the linear programming problem*

$$\min_x c^\top x, \ \ s.t.\ Ax = b, x \geq 0,$$

and let $(\tilde{x}_\mu, \tilde{y}_\mu, \tilde{s}_\mu) \in \mathcal{N}_2(\theta)$ *with average complementarity product* $\tilde{x}_\mu^\top \tilde{s}_\mu / n = \mu$. *Then there are constants* $C_x, C_s > 0$, *only dependent on the problem data and* μ, *but not on* θ, *such that*

$$\begin{aligned} \|\tilde{x}_\mu - x_\mu\|_\infty &\leq C_x(\mu)\frac{\theta}{1-\theta}, \\ \|\tilde{s}_\mu - s_\mu\|_\infty &\leq C_s(\mu)\frac{\theta}{1-\theta}, \\ \|\tilde{y}_\mu - y_\mu\|_\infty &\leq C_x(\mu)\frac{\theta}{1-\theta}. \end{aligned}$$

Proof. Let $\bar{\mu} \in \mathcal{R}_+^n$ and $(x(\bar{\mu}), y(\bar{\mu}), s(\bar{\mu}))$ be the unique solution to

$$\begin{aligned} A^\top y(\bar{\mu}) + s(\bar{\mu}) &= c \\ Ax(\bar{\mu}) &= b \\ S(\bar{\mu})X(\bar{\mu})e &= \bar{\mu} \\ s(\bar{\mu}), x(\bar{\mu}) &> 0 \end{aligned} \tag{19}$$

then we have $(x_\mu, y_\mu, s_\mu) = (x(\mu e), y(\mu e), s(\mu e))$ and there is a $\tilde{\mu} \in \mathcal{R}_+^n$, such that

$$\|\mu e - \tilde{\mu}\|_2 \leq \theta\mu, \quad e^\top \tilde{\mu}/n = \mu \tag{20}$$

with

$$(\tilde{x}_\mu, \tilde{y}_\mu, \tilde{s}_\mu) = (x(\tilde{\mu}), y(\tilde{\mu}), s(\tilde{\mu})).$$

Differentiating (19) with respect to a component $\bar{\mu}_j$ of $\bar{\mu}$ gives

$$\begin{aligned} A^\top \tfrac{\mathrm{d}y(\bar{\mu})}{\mathrm{d}\bar{\mu}_j} + \tfrac{\mathrm{d}s(\bar{\mu})}{\mathrm{d}\bar{\mu}_j} &= 0 \\ A\tfrac{\mathrm{d}x(\bar{\mu})}{\mathrm{d}\bar{\mu}_j} &= 0 \\ S(\bar{\mu})\tfrac{\mathrm{d}x(\bar{\mu})}{\mathrm{d}\bar{\mu}_j} X(\bar{\mu})\tfrac{\mathrm{d}s(\bar{\mu})}{\mathrm{d}\bar{\mu}_j} &= e_j, \end{aligned}$$

which we can solve for $\frac{\mathrm{d}x(\bar{\mu})}{\mathrm{d}\bar{\mu}_j}, \frac{\mathrm{d}y(\bar{\mu})}{\mathrm{d}\bar{\mu}_j}, \frac{\mathrm{d}y(\bar{\mu})}{\mathrm{d}\bar{\mu}_j}$ to get

$$\frac{\mathrm{d}y(\bar{\mu})}{\mathrm{d}\bar{\mu}_j} = -(AXS^{-1}A^\top)^{-1}AS^{-1}e_j, \tag{21a}$$

$$\frac{\mathrm{d}s(\bar{\mu})}{\mathrm{d}\bar{\mu}_j} = A^\top(AXS^{-1}A^\top)^{-1}AS^{-1}e_j, \tag{21b}$$

$$\frac{\mathrm{d}x(\bar{\mu})}{\mathrm{d}\bar{\mu}_j} = (I - S^{-1}XA^\top(AXS^{-1}A^\top)^{-1}A)S^{-1}e_j. \tag{21c}$$

For any $(x, y, s) \in \mathcal{N}_2(\theta)$ we have

$$(1-\theta)\mu \leq x_j s_j \leq (1+\theta)\mu$$

and hence $s_j^{-1} \leq \frac{1}{(1-\theta)\mu} x_j$ which yields

$$\|S^{-1}\|_\infty \leq \frac{1}{(1-\theta)\mu}\|X\|_\infty.$$

Further, from [18, Theorem 3.1] we have the relations

$$|x_j| \leq 2C(d)[C(d)\|d\| + \mu n], \qquad |s_j| \leq 2C(d)[C(d)\|d\| + \mu n],$$

which together with (21c) give the bound

$$\|\frac{\mathrm{d}x(\bar{\mu})}{\mathrm{d}\bar{\mu}_j}\|_\infty \leq \frac{2}{(1-\theta)\mu}(1+\chi(A))\|A\|_\infty)C(d)[C(d)\|d\| + \mu n].$$

For a bound on $\|\frac{\mathrm{d}s(\bar{\mu})}{\mathrm{d}\bar{\mu}_j}\|_\infty$ we can rewrite (21b) as

$$\frac{\mathrm{d}s(\bar{\mu})}{\mathrm{d}\bar{\mu}_j} = (X^{-1}S)XS^{-1}A^\top(AXS^{-1}A^\top)^{-1}AS^{-1}e_j,$$

and using $s_j/x_j = s_j^2/(x_j s_j) \leq \frac{1}{\mu(1-\theta)}|s_j|^2$ we obtain

$$\begin{array}{rl} \|\frac{\mathrm{d}s(\bar{\mu})}{\mathrm{d}\bar{\mu}_j}\|_\infty & \leq \frac{1}{\mu^2(1-\theta)^2}\chi(A)\|A\|_\infty\|X\|_\infty\|S\|_\infty^2 \\ & \leq \frac{8}{\mu^2(1-\theta)^2}\chi(A)\|A\|_\infty C(d)^3[C(d)\|d\| + \mu n]^3. \end{array} \tag{22}$$

Finally, for a bound on $\|\frac{\mathrm{d}y(\bar{\mu})}{\mathrm{d}\bar{\mu}_j}\|_\infty$ we can rewrite (21a) as

$$\frac{\mathrm{d}y(\bar{\mu})}{\mathrm{d}\bar{\mu}_j} = -(AXS^{-1}A^\top)^{-1}AS^{-1}X(X^{-1}e_j),$$

which yields

$$\|\frac{\mathrm{d}y(\bar{\mu})}{\mathrm{d}\bar{\mu}_j}\|_\infty \leq \chi(A)\|X^{-1}\| \leq \frac{2}{\mu(1-\theta)}\chi(A)C(d)[C(d)\|d\| + \mu n].$$

By defining

$$\begin{array}{rcl} C_x = C_x(\mu) & = & \frac{2}{\mu}(1+\chi(A))\|A\|_\infty C(d)[C(d)\|d\| + \mu n] \\ C_s = C_s(\mu) & = & \frac{8}{\mu^2}\chi(A)\|A\|_\infty C(d)^3[C(d)\|d\| + \mu n]^3 \end{array}$$

and since $1 \leq 1/(1-\theta)$ we have

$$\|\frac{\mathrm{d}x(\bar{\mu})}{\mathrm{d}\bar{\mu}_j}\|_\infty \leq \frac{1}{1-\theta}C_x, \qquad \|\frac{\mathrm{d}s(\bar{\mu})}{\mathrm{d}\bar{\mu}_j}\|_\infty \leq \frac{1}{(1-\theta)^2}C_s, \qquad \|\frac{\mathrm{d}y(\bar{\mu})}{\mathrm{d}\bar{\mu}_j}\|_\infty \leq \frac{1}{1-\theta}C_x.$$

Together with (20) we get

$$\begin{array}{rcl} \|\tilde{x}_\mu - x_\mu\|_\infty & \leq & \frac{1}{1-\theta}C_x\|\mu e - \tilde{\mu}\|_\infty \leq \frac{\theta}{1-\theta}C_x(\mu)\mu \\ \|\tilde{s}_\mu - s_\mu\|_\infty & \leq & \frac{1}{(1-\theta)^2}C_s\|\mu e - \tilde{\mu}\|_\infty \leq \frac{\theta}{(1-\theta)^2}C_s(\mu)\mu \\ \|\tilde{y}_\mu - y_\mu\|_\infty & \leq & \frac{1}{1-\theta}C_x\|\mu e - \tilde{\mu}\|_\infty \leq \frac{\theta}{1-\theta}C_x(\mu)\mu \end{array}$$

□

Lemma 6. *For $\theta \in (0,1)$, let $(\tilde{x}_\mu^R, \tilde{y}_\mu^R, \tilde{\lambda}_\mu^R, \tilde{s}_\mu^R, \tilde{z}_\mu^R) \in \mathcal{N}_2^R(\theta)$. Then there is a $C_3 > 0$ independent of θ such that*

$$\begin{array}{rcl} \|\tilde{x}_\mu^R - x_\mu(\mathcal{T}^R)\|_\infty & \leq & C_3\frac{\theta}{1-\theta}, \\ \|\tilde{\lambda}_\mu^R - \lambda_\mu(\mathcal{T}^R)\|_\infty & \leq & C_3\frac{\theta}{1-\theta}, \\ \|\tilde{s}_\mu^R - s_\mu(\mathcal{T}^R)\|_\infty & \leq & C_3\frac{\theta}{(1-\theta)^2}. \end{array}$$

Proof. This is an immediate consequence of Lemma 5. □

From the previous lemma we get that we can bound the difference in the primal–dual first-stage decisions (x, s) of the true μ-center of the full problem $(x_\mu(\mathcal{T}), s_\mu(\mathcal{T}))$ to the calculated approximate μ-center for the reduced problem $(\tilde{x}^R_\mu, \tilde{s}^R_\mu)$ by

$$\begin{aligned}\|x_\mu(\mathcal{T}) - \tilde{x}^R_\mu\|_\infty &\leq \frac{2}{\sqrt{\mu}}\overline{B}\sqrt{L_Q}\sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} + C_x(\mu)\mu\frac{\theta}{1-\theta}\\ \|s_\mu(\mathcal{T}) - \tilde{s}^R_\mu\|_\infty &\leq \frac{2}{\sqrt{\mu}^3}\overline{B}^3\sqrt{L_Q}\sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} + C_s(\mu)\mu\frac{\theta}{(1-\theta)^2}\end{aligned}$$

In the second step of the algorithm we will not find the exact μ-center for all subproblems $P_i(\tilde{x}^R_\mu)$, but rather find points

$$(\tilde{y}_i, \tilde{\lambda}_i, \tilde{s}_i) \in \mathcal{N}_2^{(i)}(\theta).$$

Again we need a bound on the implied error in the dual components λ_i. According to Lemma 6 there is a $C_3 > 0$ such that

$$\|\tilde{\lambda}_i - \lambda_i\|_\infty \leq C_3\frac{\theta}{1-\theta}.$$

This affects the bound on $\|\Delta c\|$ in the proof of Theorem 4.

Theorem 7. *Let $\tilde{w}$ be the warm-start point for problem $P(\mathcal{T})$ obtained by following the above algorithm using a centrality measure of $\tilde{\theta}$, starting from the reduced tree $\mathcal{T}^R$, using a target μ-value of $\bar{\mu}$. If we choose $W_1(\mathcal{T}, \mathcal{T}^R)$ and $\tilde{\theta}$ small enough then the warmstart is successful, that is the Yıldırım -Wright [24] Weighted Least Squares step Δw from $\tilde{w}$ is feasible and leads to $\tilde{w} + \Delta w \in \mathcal{N}_2^{\mathcal{T}}(\theta)$.*

Proof. From (17) and (18) we have that

$$\|\tilde{X}\tilde{S}e - \bar{\mu}e\|_2 \leq \tilde{\theta}\bar{\mu}, \qquad \|\tilde{Y}_i\tilde{Z}_ie - \bar{\mu}e\|_2 \leq \tilde{\theta}\bar{\mu}$$

and therefore

$$\begin{aligned}\|(\tilde{X}, \tilde{Y}_1, \ldots, \tilde{Y}_n)(\tilde{S}, \tilde{Z}_1, \ldots, \tilde{Z}_n)e - \bar{\mu}e\|_2^2 &= \|\tilde{X}\tilde{S}e - \bar{\mu}e\|_2^2 + \sum_i \|\tilde{Y}_i\tilde{Z}_ie - \bar{\mu}e\|_2^2\\ &\leq (|\mathcal{T}| + 1)\tilde{\theta}^2\bar{\mu}^2. \qquad (23)\end{aligned}$$

As in the proof to Theorem 4 let

$$\tilde{c} = \sum_i T^\top\tilde{\lambda}_i + \tilde{s}$$

and consider the problem instance $P(\tilde{d})$ obtained from $\mathcal{P}(\mathcal{T})$ by replacing the first-stage cost c with $\tilde{c}$. Then by construction the warm-start point $\tilde{w}$ is primal–dual feasible for $P(\tilde{d})$ and due to (23) satisfies

$$\tilde{w} \in \mathcal{N}_2^{\tilde{d}}(\sqrt{|\mathcal{T}+1|}\tilde{\theta}).$$

We will analyse the warmstart as a warmstart attempt for the (perturbed) problem $P(\mathcal{T})$ starting from a point in the $\mathcal{N}_2$-neighbourhood for problem $P(\tilde{d})$. The change in problem data is

$$\Delta d = (\Delta A, \Delta b, \Delta c) = (0, 0, \Delta c)$$

with

$$\Delta c = c - \tilde{c} = \sum_i T^\top \lambda_{\mu,i} + s_\mu - (\sum_i T^\top \tilde{\lambda}_i + \tilde{s}) = \sum_i T^\top (\lambda_{\mu,i} - \tilde{\lambda}_i) + (s_\mu - \tilde{s}).$$

Using the bounds from Lemma 3 and Lemma 6 we have

$$\begin{aligned}
\|\Delta c\| &= \|\sum_i T^\top (\lambda_{\mu,i} - \bar{\lambda}_i + \bar{\lambda}_i - \tilde{\lambda}_i) + (s_\mu - \bar{s} + \bar{s} - \tilde{s})\| \\
&\leq \|T\||\mathcal{T}|(\|\lambda_{\mu,i} - \bar{\lambda}_i\| + \|\bar{\lambda}_i - \tilde{\lambda}_i\|) + \|s_\mu - \bar{s}\| + \|\bar{s} - \tilde{s}\| \\
&\leq \|T\||\mathcal{T}| \left(C_\lambda \|x_\mu - \tilde{x}\| + C_3 \frac{\tilde{\theta}}{1-\tilde{\theta}} \right) + \|s_\mu - \bar{s}\| + C_3 \frac{\tilde{\theta}}{1-\tilde{\theta}}.
\end{aligned}$$

Moreover, using the bounds from Lemma 2 we get

$$\begin{aligned}
\|\Delta c\| &\leq \|T\||\mathcal{T}| \left(C_\lambda \frac{2}{\sqrt{\mu}} \sqrt{L_Q} \bar{B} \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} + C_3 \frac{\tilde{\theta}}{1-\tilde{\theta}} \right) \\
&\quad + \frac{2}{\sqrt{\mu^3}} \sqrt{L_Q} \bar{B}^3 \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} + C_3 \frac{\tilde{\theta}}{1-\tilde{\theta}} \\
&= \left(\|T\||\mathcal{T}| C_\lambda(\mu) + \frac{\bar{B}^2}{\mu} \right) \frac{2}{\sqrt{\mu}} \sqrt{L_Q} \bar{B} \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} + (\|T\||\mathcal{T}| C_\lambda(\mu) + 1) C_3 \frac{\tilde{\theta}}{1-\tilde{\theta}} \\
&\leq C_4(\mu) \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} + C_5 \frac{\tilde{\theta}}{1-\tilde{\theta}} \qquad (24)
\end{aligned}$$

where

$$C_4(\mu) = \left(\|T\||\mathcal{T}| C_\lambda(\mu) + \frac{\bar{B}^2}{\mu} \right) \frac{2}{\sqrt{\mu}} \bar{B} \sqrt{L_Q}, \qquad C_5 = (\|T\||\mathcal{T}| C_\lambda(\mu) + 1) C_3.$$

Proposition 4.2 of [24] can now be applied with $\theta_0 = \tilde{\theta}\sqrt{|\mathcal{T}|+1}$ and $\xi = \frac{1}{2}(\theta - \tilde{\theta}\sqrt{|\mathcal{T}|+1})$ from which we get

$$\theta - \theta_0 - \xi = \frac{1}{2}(\theta - \tilde{\theta}\sqrt{|\mathcal{T}|+1})$$

and therefore the conditions for a successful warmstart are

$$\|\Delta c\|_\infty \leq \frac{\theta - \tilde{\theta}\sqrt{|\mathcal{T}|+1}}{2(2n+1)\overline{C(d)}} \overline{\|d\|}, \qquad \mu \geq \frac{8\overline{C(d)}^2}{\theta - \tilde{\theta}\sqrt{|\mathcal{T}|+1}} \|\Delta c\|_\infty$$

which can be combined to obtain

$$\|\Delta c\|_\infty \leq \frac{\theta - \tilde{\theta}\sqrt{|\mathcal{T}|+1}}{2\overline{C(d)}} \min \left\{ \frac{\overline{\|d\|}}{2n+1}, \frac{\mu}{4\overline{C(d)}} \right\}. \qquad (25)$$

Combining (24) and (25) we get as the condition for a successful warmstart

$$C_4(\mu) \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} + C_5 \frac{\tilde{\theta}}{1-\tilde{\theta}} \leq \frac{\theta - \tilde{\theta}\sqrt{|\mathcal{T}|+1}}{2\overline{C(d)}} \min \left\{ \frac{\overline{\|d\|}}{2n+1}, \frac{\mu}{4\overline{C(d)}} \right\}$$

which can be satisfied by keeping $W_1(\mathcal{T}, \mathcal{T}^R)$ and $\tilde{\theta}$ small enough. □

5 Implementation and numerical results

In this section we present the numerical results of the implementation of the proposed decomposition warm-start scheme within the interior point code OOPS.

The setup is the following. Given a 2-stage stochastic programming problem with a given scenario set $\mathcal{T}$, we first generate a reduced tree $\mathcal{T}^R$ with a specified number of scenarios by eliminating scenarios using the scenario reduction technique of [13]. The reduced problem is solved until the first iterate for which $\mu^{(k)} < \bar{\mu}$. If this point is not primal-dual feasible, additional pure centering iterations with a target of $\mu = \bar{\mu}$ are performed until we obtain a primal-dual feasible point. The x-components of this point are used to set-up scenario subproblems $P_{i,\mu}(x)$ which again are solved to obtain a primal-dual feasible point corresponding to $\mu \approx \bar{\mu}$ following the procedure above. The algorithm is to some extent sensitive to the correct choice of the target $\bar{\mu}$. We report results for the best $\bar{\mu}$ value found in our limited experiments (using trial values of $\bar{\mu} = 10^k, k = -4, \ldots, 1$). This is of course slightly unsatisfactory. Indications are however that the best $\bar{\mu}$ value is fairly constant within a problem class.

All computations were performed on a Linux PC with a 3.0GHz Intel Core 2 processor and 3GB of RAM. Final convergence tolerance was to reduce the relative primal-dual gap below 10^{-7}.

We concentrate our tests on the two-stage problems summarized in Table 1. The first block of problems are from the collection of Ariyawansa and Felt [2], '`sslp`' and '`dcap`' are from the SIPLIB collection of Ahmed [1] and `storm` is from the POSTS collection [4]. We compare solution number of iterations until converged from the crash-start point and total running time (for both crash-start generation and solution phase) for the suggested algorithm against a cold-start using Mehrotra's starting point. Results are summarized in Table 2. As can be seen the decomposition-based crash start offers significant savings both in terms of iterations and time. Savings average about 60% in terms of iterations and 30% in terms of time and are more pronounced for larger problems, in many cases achieving a halving of total solution time.

Finally we have evaluated the algorithm on some multistage problems from the literature. For multistage problems the proposed decomposition based crash-start will decompose the problem at the second stage and then solve each subtree originating from a second stage node separately. This approach is in effect treating a multistage problem as a two stage problem with a large core matrix. The reduced scenario tree $\mathcal{T}^R$ is obtained by choosing a single scenario originating from every second stage node. We choose the one resulting in the minimal Wasserstein distance from the remaining nodes in this subtree.

Our selecion of multistage problems together with the problem characteristics is summarised in Table 3. Results are given in Table 4: as before we are comparing number of iterations for the warmstarted problem and total solution time to number of iterations and solution time for a cold-started IPM using Mehrotra's starting point. Again we can see that the crash-start saves a significant number of iterations, although this is not translated into time savings in the same way as for two-stage problems.

Problem	scenarios	rows	columns	nonzeros
AIRL1	25	152	306	706
AIRL2	25	152	306	706
AIRL3	676	4,058	8,118	18,934
cargo-4node32	32	2,382	6,396	15,656
cargo-4node64	64	4,750	12,732	31,016
cargo-4node128	128	9,486	25,404	61,736
cargo-4node256	256	18,958	50,748	123,176
cargo-4node512	512	37,902	101,436	246,056
cargo-4node1024	1024	75,790	202,812	491,816
cargo-4node2048	2048	151,566	405,564	983,336
cargo-4node4096	4096	303,118	811,068	1,966,376
cargo-4node8192	8196	606,222	1,622,076	3,932,456
cargo-4node16384	16384	1,212,430	3,244,092	7,864,616
asset1	100	505	1,313	2,621
asset2	37500	187,505	487,513	975,021
env.1200	1200	57,648	102,085	220,972
env.1875	1875	90,048	159,460	345,172
env.3780	3780	181,488	321,385	695,692
env.5292	5292	254,064	449,905	973,900
env.lrge	8232	395,184	699,805	1,514,860
stocfor2	64	6,543	9,237	29,985
sslp_10_50_100	100	6,001	52,011	101,911
sslp_10_50_500	500	30,001	260,011	509,511
sslp_10_50_1000	1000	60,001	520,011	1,019,011
dcap233_500	500	7,506	16,518	31,518
dcap243_500	500	9,006	21,018	39,018
storm27	27	14,441	37,485	94,274
storm125	125	66,185	172,431	433,256
storm1000	1000	528,185	1,377,306	3,459,881

Table 1: Characteristics of 2-stage problems

Problem	$\|\mathcal{T}\|$	$\|\mathcal{T}^R\|$	cold		dec		
			iter	time	iter	time	$\bar{\mu}$
AIRL1	25	10	12	0.08	11	0.16	1.0
AIRL2	25	10	10	0.07	5	0.16	1.0
AIRL3	676	60	13	2.8	5	3.5	1.0
cargo-4node32	32	6	21	2.6	9	2.5	1.0
cargo-4node64	64	10	26	6.4	10	4.9	1.0
cargo-4node128	128	10	26	13.2	9	8.2	0.1
cargo-4node256	256	10	23	23.8	8	11.2	0.01
cargo-4node512	512	10	30	62.9	7	20.6	0.01
cargo-4node1024	1024	50	30	123	9	52	0.01
cargo-4node2048	2048	200	34	297	11	134	0.01
cargo-4node4096	4096	200	41	710	12	277	0.01
cargo-4node8192	8192	400	39	1118	12	487	0.01
cargo-4node16384	16384	800	39	2349	12	708	0.01
asset1	100	10	12	0.3	5	0.3	0.01
env.1200	1200	50	32	61	12	32	0.01
env.1875	1875	50	32	102	8	41	0.001
env.3780	3780	50	34	218	10	97	0.001
env.5292	5292	200	34	313	10	147	0.001
env.lrge	8232	200	39	577	11	258	0.001
stocfor2	64	2	21	3.9	29	4.8	0.01
sslp_10_50_100	100	10	26	12.6	15	12.3	0.001
sslp_10_50_500	500	50	51	130	21	81	0.0001
sslp_10_50_1000	1000	50	56	301	22	163	0.0001
dcap233_500	500	50	22	6.5	7	4.3	0.01
dcap243_500	500	50	23	7.8	8	5.3	0.001
storm27	27	5	88	61	12	18.8	0.1
storm125	125	10	96	306	8	61	1.0
storm1000	1000	100	103	2351	22	771	1.0

Table 2: Results for 2-stage problems.

6 Conclusions and future research

In this paper we proposed a technique to generate a crash-start point for interior point methods applied to stochastic programming problems, based on solving a the problem on a reduced tree and a sequence of scenario subproblems. The scheme can be applied to two-stage and multi-stage problems.

We show that this approach leads to an advanced iterate from which an Interior Point method can be warmstarted succesfully, providing the reduced tree is a good approximation to the full tree in an appropriate measure. Numerical results demonstrate that the approach results in significant time savings, especially on larger problems.

For problems with a large number of second stage nodes, solving all scenario subproblems can be a significant cost. In our implementation the scenario subproblems are solved independently,

Problem	stages	scenarios	rows	columns	nonzeros
fxm3-6	3	36	6,200	12,628	57,722
fxm3-16	3	256	41,340	85,575	392,252
fxm4-6	4	216	22,400	47,185	265,442
stocfor3	7	512	16,675	23,541	76,473
pltexpA3-16	3	256	28,350	74,172	150,801
pltexpA4-6	4	216	26,894	70,364	143,059
pltexpA5-6	5	1,296	161,678	422,876	859,747
swing8-4	8	65,536	262,142	349,522	786,422
watson10-128	8	128	26,237	49,664	129,159
watson10-512	10	512	67,069	129,024	351,751
watson10-1024	10	1024	134,127	258,032	703,473

Table 3: Characteristics of multistage test problems.

Problem	cold		dec		
	iter	time	iter	time	$\bar{\mu}$
fxm3-6	26	7.8	6	6.8	0.1
fxm3-16	30	60.8	9	45.5	0.1
fxm4-6	25	31.2	8	29.0	0.1
stocfor3	32	22.9	15	19.5	0.1
pltexpA3-16	35	36.3	12	29.4	0.00001
pltexpA4-6	42	46.0	10	28.3	0.0001
pltexpA5-6	75	528	39	401	0.0001
swing8-4	28	153	10	120	0.01
watson10-128	55	48.3	10	37.2	0.001
watson10-512	80	193	7	148	0.0001
watson10-1024	94	457	12	323	0.0001

Table 4: Number of iterations and time in seconds to solve the complete problem to optimality for different starting points.

using Mehrotras starting point. Clearly since the scenario subproblems are usually closely related, they could be warmstarted as well, leading to further efficiency gains. We leave this for future research.

References

[1] S. Ahmed, *SIPLIB: A stochastic integer programming test problem library.* http://www2.isye.gatech.edu/~sahmed/siplib/, 2004.

[2] K. A. Ariyawansa and A. J. Felt, *On a new collection of stochastic linear programming test problems*, INFORMS Journal on Computing, 16 (2004), pp. 291–299.

[3] J. R. Birge, *Decomposition and partitioning methods for multistage stochastic linear programs*, Oper Res, 33 (1985), pp. 989–1007.

[4] J. R. Birge and D. Holmes, *A portable stochastic programming test set (POSTS).* http://users.iems.northwestern.edu/~jrbirge/html/dholmes/post.html.

[5] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer-Verlag, New York, 1997.

[6] J. Blomvall and P. O. Lindberg, *A Riccati-based primal interior point solver for multistage stochastic programming*, Eur J Oper Res, 143 (2002), pp. 452–461.

[7] M. Colombo, J. Gondzio, and A. Grothey, *A warm-start approach for large-scale stochastic linear programs*, Math Program, (2009). published online.

[8] I. I. Dikin, *On the speed of an iterative process*, Upravlaemye Sistemy, 12 (1974), pp. 54–60.

[9] J. Dupačová, N. Gröwe-Kuska, and W. Römisch, *Scenario reduction in stochastic programming*, Math Program, 95 (2003), pp. 493–511.

[10] J. Gondzio and A. Grothey, *Reoptimization with the primal-dual interior point method*, SIAM Journal on Optimization, 13 (2003), pp. 842–864.

[11] ——, *A new unblocking technique to warmstart interior point methods based on sensitivity analysis*, SIAM J Opimiz, 19 (2008), pp. 1184–1210.

[12] J. Gondzio and R. Sarkissian, *Parallel interior point solver for structured linear programs*, Math Program, 96 (2003), pp. 561–584.

[13] W. R. H. Heitsch, *Scenario reduction algorithms in stochastic programming*, Computational Optimization and Applications, 24 (2003), pp. 187–206.

[14] A. L. Hipolito, *A weighted least squares study of robustness in interior point linear programming*, Computational Optimization and Applications, 2 (1993), pp. 29–46.

[15] P. Kall and S. W. Wallace, *Stochastic Programming*, John Wiley & Sons, Chichester, 1994.

[16] S. Mehrotra, *On the implementation of a primal–dual interior point method*, SIAM Journal on Optimization, 2 (1992), pp. 575–601.

[17] J. Mulvey and A. Ruszczyński, *A new scenario decomposition method for large scale stochastic optimization*, Oper Res, 43 (1995), pp. 477–490.

[18] M. A. Nunez and R. M. Freund, *Condition measures and properties of the central trajectory*, Math Program, 83 (1998), pp. 1–28.

[19] W. Römisch and R. Schultz, *Stability analysis for stochastic programming*, Annals of Operations Research, 30 (1991), pp. 241–266.

[20] R. Schultz, *Some aspects of stability in stochstic programming*, Ann Oper Res, 100 (2000), pp. 55–84.

[21] M. Steinbach, *Hierarchical sparsity in multistage convex stochastic programs*, in Stochastic Optimization: Algorithms and Applications, S. Uryasev and P. M. Pardalos, eds., Kluwer Academic Publishers, 2000, pp. 363–388.

[22] J. Sun and X. Liu, *Scenario formulation of stochastic linear programs and the homogeneous self-dual interior point method*, INFORMS Journal on Computing, 18 (2006), pp. 444–454.

[23] S. J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, 1997.

[24] E. A. Yildirim and S. J. Wright, *Warm-start strategies in interior-point methods for linear programming*, SIAM Journal on Optimization, 12 (2002), pp. 782–810.