

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

Abstract

In this paper we propose a warm-start technique for interior point methods applicable to multi-stage stochastic linear programming problems. The main idea is to generate an initial point by decomposing the problem at the second stage and using an approximate solution 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

In a previous work [4], the authors have introduced a warm-start approach for stochastic programming based on solving a reduced problem which contains only a trivial fraction of the scenarios. This exploited the fact that the stochastic programming framework embeds a structured problem, which can be easily reduced in dimension while keeping intact the “essence” of the structure. While the concept of *structure* has been exploited in several fashions in decomposition-based approaches (Dantzig–Wolfe and Benders’ decomposition) and in the linear algebra [3, 7, 14], it was not employed in an algorithmic way to generate a starting point for an interior point method. Moreover, that work provided measurable evidence that interior point methods can be efficiently warm-started.

The possibility of warm-starting a large-scale problem is an important feature in a solution method. As the problem dimensions increase, also computation time and memory requirements

*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/>

grow. Warm-starting, that is, providing and employing an advanced starting point obtained by solving a related problem, proves to be a valuable way to cut the number of iterations necessary to find the optimal solution. Provided that finding the warm-start point is relatively cheap, then this usually translates in computational time savings.

The approach described in [4] is efficient albeit crude. One of its drawbacks is that the initial point, built from a reduced scenario tree, is used as is for parts of the tree that may be far away. 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. The paper [4] describes the conditions for a successful warm-start. They can be summarised in the requirement of the reduced tree to be close to the original tree.

Therefore, it becomes essential to deal with reduced trees that, being closer to the original event tree, generate a warm-start point that satisfies the theoretical conditions analysed. This observation provides the motivation of the current investigation. The idea is that the warm-start point generated from a reduced tree can be tailored to a specific subtree by solving a subproblem rooted at the topmost node of the subtree. This approach has clear connections with a decomposition approach. In this paper we describe how the decomposition idea can be used to generate a warm-start point and we provide numerical evidence of its advantages.

The paper is structured as follows. In Section 2 we first review the warm-start approach of [4] in the context of stochastic programming. In Section 3 we introduce the decomposition idea, 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 [2, 8] 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

$$\begin{aligned}
 \min_x \quad & c^\top x + \mathbb{E}_\xi[Q(x, \xi)] & \text{where} \quad & Q(x, \xi) = \min_y \quad c(\xi)^\top y(\xi) \\
 \text{s.t.} \quad & W_1 x = h_1 & & \text{s.t.} \quad T x + W(\xi)y(\xi) = h(\xi) \\
 & x \geq 0 & & y(\xi) \geq 0,
 \end{aligned} \tag{1}$$

where the random variable ξ captures the uncertainty. 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

$$\begin{aligned} \min \quad & c^\top x + \sum_{i \in \mathcal{T}} \pi_i c_i^\top y_i \\ & Tx + W_i y_i = h_i, \quad i \in \mathcal{T} \\ & x, y_i \geq 0. \end{aligned} \quad (2)$$

We can introduce a multi-stage decision structure by considering several stages of recourse decisions (x, y^1, \dots, y^T) and a tree of scenarios. For the purposes of this paper we will assume that the stages $2, \dots, T$ are represented as one, that is y_i in (2) denotes the vector $y_i = (y_i^1, \dots, y_i^T)$ 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 ([1, 10, 15]). 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 the algorithm to 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

$$\min c^\top x \quad \text{s.t.} \quad Ax = b, \quad x \geq 0, \quad (3)$$

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 (3) 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

$$\min c^\top x - \mu \sum \ln x_i \quad \text{s.t.} \quad Ax = b, \quad (4)$$

instead of (3). Problem (4) is a family of strictly convex problems, parameterised by $\mu > 0$, whose unique solution approaches the solution to the original problem as $\mu \rightarrow 0$. The trajectory of solutions to (4) for different values of μ is the *central path*. Interior point methods apply a damped Newton method to the optimality conditions of (4), yielding the direction-finding problem

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^\top & I \\ Z & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} b - Ax \\ c - A^\top y - z \\ -XZe + \mu e \end{bmatrix} = \begin{bmatrix} \xi_b \\ \xi_c \\ \xi_\mu \end{bmatrix}, \quad (5)$$

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 [16] globalise the Newton iteration by keeping the iterates in a neighbourhood of the central path, thus follow it in approaching the optimal solution. 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 [9], 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 $(x, z) > 0$. Mehrotra’s starting point heuristic does not exploit the structural information available in the problem. We argue that taking structure into account when choosing the first iterate can lead to a faster identification of the correct active set, thus speeding up the solution process.

In the warm-start approach of [4], from the event tree \mathcal{T} associated to the problem, we build a reduced tree, \mathcal{T}^R , by choosing just some of the available scenarios. We find an approximate solution to the deterministic equivalent corresponding to \mathcal{T}^R and use it to construct a warm-start iterate for the complete problem, which is then solved to optimality. The advantage of the scheme 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. We use an interior point method to find a sufficiently advanced primal–dual feasible point for the reduced problem: therefore, rather than achieving optimality, we stop at an iterate corresponding to a few digits of accuracy in the solution.

The scheme of [4], as well as the approach discussed in this paper, generate an initial point from a reduced set of scenarios. To understand why the warm-start approach of [4] can be very successful even if the reduced tree contains only 2 of the original set of scenarios, we studied the active set of some of the problems. If, for each scenario considered, the active set is similar, then even by using a fraction of the scenarios it is possible to generate correctly the active set for all scenarios.

To investigate this claim, we considered some two-stage problems from a standard set of test problems. We solved the problems to optimality, then extracted the active set information, and obtained the pictures shown in Figure 1. Each horizontal line corresponds to a scenario: for each variable within a scenario, a yellow dot was printed if the variable was active, a black dot otherwise. Thus, looking vertically, we can spot the difference in active set between each scenario.

It is striking to see the great similarity among all of them for the problems considered. Most of the scenarios lead to the same active set, and only a trivial fraction of variables display some difference. This seems to validate what we have seen earlier: we can obtain a correct localisation of the optimal active set by considering only part of the scenario tree.

3 Decomposition scheme

We implemented a new way of producing a warm-start iterate for stochastic linear programming. The underlying idea recalls that of decomposition schemes such as Benders’ decomposition.



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

From formulation (1), given a first-stage decision x , we define the second-stage problem as

$$Q(x, \xi) = \min\{c(\xi)^\top y : W(\xi)y = h(\xi) - Tx, y \geq 0\}.$$

We remark that since in general we are interested in multi-stage programs, we interpret W above as a structured matrix that may correspond to more than one stage. This allows us to write a generic stochastic linear programming problem with recourse in the form

$$\min \mathbb{E}_\xi\{c^\top x + Q(x, \xi)\} \quad \text{s.t.} \quad W_1x = h_1, x \geq 0.$$

It is evident that the presence of the first-stage decision variables creates connections with the blocks corresponding to the second-stage nodes. This relationship is what guarantees the non-anticipativity of the decisions, thus ensuring that the decisions at one stage do not depend from the stochastic realisations at the next stages. On the other hand, if such connections were absent, then the problem would decompose in a number of subproblems, each one corresponding to a second-stage node. Under this assumption, the variables in the subproblems would not display any dependence, and each subproblem could be solved independently.

Therefore, after fixing x , the linking terms can be eliminated to obtain a series of smaller, independent linear programs. In terms of the deterministic equivalent formulation, problem (2) becomes

$$\min\{c^\top x + \sum Q_i(x) : W_1x = h_1, x \geq 0\},$$

where we define the scenario subproblems

$$Q_i(x) = \min\{\pi_i c_i^\top y_i : W y_i = h_i - T_i x, y_i \geq 0\}, \quad (P_i(x))$$

and each subproblem is rooted at a second-stage node.

The dual solution of problem $(P_i(x))$ can be used to find subgradients for $Q_i(x)$, which then allow to construct a convex estimate of $Q(x)$. Thus, the minimization of $Q(x)$ is recast into a minimization of its piecewise linear approximation. This approach is usually further refined by considering quadratic regularisation terms.

The basic idea of Benders' decomposition is that the first-stage problem, referred to as the master problem, provides some values for the first-stage decision x : these are seen as *query points*, which are passed to the subproblems. Each subproblem provides a feedback on the query point: if it is feasible, and hence it allows to find an optimal solution, then the subproblem will return the achievable optimum $Q_i(x)$; otherwise, it will produce a constraint that, added to the master problem, will exclude the current query point.

3.1 A decomposition-based warm-start

The main difficulty in decomposition methods is to identify the optimal values for the first stage decisions. In Benders' decomposition this is achieved by solving master problems involving a piecewise linear approximation of the recourse function $Q(x)$, but the process may be slow. Once the optimal first stage decisions are known, the rest of the solution can be obtained by solving each scenario subproblem in a single iteration. In the proposed decomposition-based warm-start scheme we obtain an estimate for the optimal first stage decisions by solving the problem on a reduced tree. Further, since we are interested in finding a warm-start point, rather than the optimal solution, we apply the decomposition procedure to the barrier problem (4) for a suitable value of μ .

The suggested warm-start approach is summarised in the following algorithm:

Algorithm 1 Decomposition-based warm-start algorithm

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

- 1: Generate a reduced event tree $\mathcal{T}^R \subset \mathcal{T}$;
 - 2: Solve the deterministic equivalent corresponding to \mathcal{T}^R with a loose tolerance with the aim of generating some approximation \tilde{x} to the first-stage variables;
 - 3: Decompose the problem by setting up independent subproblems $P_i(\tilde{x})$ rooted at the second-stage nodes, using the current approximation \tilde{x} to substitute out the first-stage variables and solve the independent subproblems;
 - 4: Use the solution of the subproblems to initialise the corresponding blocks of the complete warm-start iterate;
 - 5: Solve the complete problem on \mathcal{T} from the warm-start point to optimality.
-

In our scheme, we obtain a value for x by solving the reduced-tree problem to loose accuracy. This problem, being of reduced dimension, is quickly solved. We then instantiate a subproblem for each second-stage node, which is then solved independently. The warm-start iterate is generated by piecing together the solutions obtained by the subproblems.

Compared to the approach of [4], 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.

4 Theoretical analysis

The aim of this section is to state conditions under which the decomposition-based warm-start procedure is successful, that is we are able to construct a primal–dual feasible and sufficiently central point by following it. 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 Nunez and Freund

[11] we use $d_i(x) = (W_i, h_i - Tx, \pi_i c_i)$ to refer to the scenario data 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 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 μ -central point $(x^R, y^R, \lambda^R, s^R, z^R) \in \mathcal{N}_2^R(\theta_0)$, where (λ^R, s^R, z^R) are the dual variables for the constraints $Tx + W_i x = h_i$, $x \geq 0$, and $y_i \geq 0$ respectively. We take the x^R, s^R 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 μ -central points

$$(\hat{y}_i, \hat{\lambda}_i, \hat{z}_i) \in \mathcal{N}_2^{P_i}(\theta_0), \quad \forall i \in \mathcal{T}.$$

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

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

where $\mathcal{N}_2^{\mathcal{T}}$ is the \mathcal{N}_2 -neighbourhood of the central path for problem $P(\mathcal{T})$. 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* [6, 12]. In an abstract probability space setting with event space and corresponding Borel σ -field (Ω, \mathcal{B}) , the Wasserstein distance of two measures μ, ν defined on this space is given by

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

In our case, where μ and ν are discrete measures implied by the full and reduced scenario trees as

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

the corresponding Wasserstein distance reduces to

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

Our analysis first deals with an idealised algorithm in which we chose 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 :

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

Moreover, we make use of the following relation from Dikin [5]

$$\chi(W) := \sup_{\Sigma \in D^+} \|\Sigma W^\top (W \Sigma W^\top)^{-1}\|_\infty < \infty, \quad (6)$$

where D^+ is the set of all diagonal matrices with strictly positive diagonal elements.

4.1 Results for exact subproblem solutions

We first investigate the dependence of the barrier equivalent of problem $(P_i(x))$

$$\begin{aligned} Q_\mu(x; \xi) = \min \quad & c^\top y - \mu \sum_j \ln y_j \\ \text{s.t.} \quad & Wy = h(\xi) - Tx \end{aligned} \quad (7)$$

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

Lemma 1. *The function $Q_\mu(x; \xi)$ is Lipschitz in the second argument, that is*

$$|Q_\mu(x; \xi) - Q_\mu(x; \bar{\xi})| \leq L_Q \|\xi - \bar{\xi}\|_2$$

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

Proof. Optimality conditions for (7) are

$$z + W^\top \lambda = c, \quad Wy + Tx = h(\xi), \quad YZe = \mu e. \quad (8)$$

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

$$W \frac{dy}{d\xi} = \frac{dh}{d\xi} \quad (9a)$$

$$\frac{dz}{d\xi} + W^\top \frac{d\lambda}{d\xi} = 0 \quad (9b)$$

$$Y \frac{dz}{d\xi} + Z \frac{dy}{d\xi} = 0. \quad (9c)$$

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

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

and hence

$$\frac{dy}{d\xi} = YZ^{-1}W^\top \frac{d\lambda}{d\xi}. \quad (10)$$

Multiplying from the left with W and substituting into (9a) yields

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

which together with (10) gives

$$\frac{dy}{d\xi} = YZ^{-1}W^\top(WYZ^{-1}W^\top)^{-1} \frac{dh}{d\xi}.$$

Now, recalling (6), we get the bound

$$\left\| \frac{dy}{d\xi} \right\|_\infty \leq \chi(W) \left\| \frac{dh}{d\xi} \right\|_\infty.$$

On the other hand, from the definition of Q_μ in (7) we have

$$\frac{dQ_\mu}{d\xi} = c^\top \frac{dy}{d\xi} - \mu Y^{-1} \frac{dy}{d\xi} = (c - z)^\top \frac{dy}{d\xi}.$$

From [11, Theorem 3.1] we have the bound

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

which together with $\|c\|_\infty \leq \|d\|$ and $\left\| \frac{dh}{d\xi} \right\|_\infty \leq H$ yields

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

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

In what follows we define

$$\eta_{\mathcal{T}}(x) := c^\top x - \mu \sum_j \ln x_j + \rho_{\mathcal{T}}(x) \quad (11)$$

where

$$\rho_{\mathcal{T}}(x) = \mathbb{E}_{\mathcal{T}}[Q_\mu(x; \xi)] = \sum_{(\pi_i, \xi_i) \in \mathcal{T}} \pi_i Q_\mu(x; \xi_i).$$

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

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

depends on the underlying scenario set \mathcal{T} . We need to assume that the set of feasible first-stage decisions x is bounded, that is

$$\|x_\mu(\mathcal{T})\|_\infty \leq \bar{B}, \quad \forall \mathcal{T}.$$

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

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

and

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

Proof. We make use of Remark 2.2 to [12, Proposition 2.1] (see also [13, Theorem 2.10]), which states that given $|h(z) - h(\tilde{z})| \leq L_h \|z - \tilde{z}\|$, we have

$$\left| \int h(z) \nu(dz) - \int h(z) \tilde{\nu}(dz) \right| \leq L_h W_1(\nu, \tilde{\nu}).$$

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. We have

$$\int Q_\mu(x; \xi) \nu(d\xi) = \mathbb{E}_T[Q_\mu(x; \xi)] = \rho_T(x),$$

likewise for $\tilde{\nu}$ and \mathcal{T}^R , and therefore

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

and hence also

$$|\eta_T(x) - \eta_{\mathcal{T}^R}(x)| \leq L_Q W_1(\mathcal{T}, \mathcal{T}^R). \quad (12)$$

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

$$\eta_{\mathcal{T}^R}(x_\mu(\mathcal{T}^R)) \leq \eta_{\mathcal{T}^R}(x_\mu(\mathcal{T})). \quad (13)$$

Bound (12) implies

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

and hence together with (13)

$$\eta_{\mathcal{T}^R}(x_\mu(\mathcal{T}^R)) \leq \eta_T(x_\mu(\mathcal{T})) + L_Q W_1(\mathcal{T}, \mathcal{T}^R). \quad (14)$$

The functions $\rho_T(x), \rho_{\mathcal{T}^R}(x)$ 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

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

Hence we get the bound

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

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

$$\eta_T(x_\mu(\mathcal{T}^R)) \geq \eta_T(x_\mu(\mathcal{T})) + \frac{1}{2} \|x_\mu(\mathcal{T}^R) - x_\mu(\mathcal{T})\|_\infty^2 \frac{\mu}{\bar{B}^2}. \quad (15)$$

Therefore, by combining (12) and (15) we get

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

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 (14) so that we have

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

After rearrangement we are left with

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

or equivalently

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

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

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

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

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

yielding the required bound. \square

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 $(x_\mu, y_\mu, \lambda_\mu, s_\mu, z_\mu)$ 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 (8). Differentiating with respect to x gives:

$$\frac{dz_i}{dx} + W^\top \frac{d\lambda_i}{dx} = 0, \quad W \frac{dy_i}{dx} = -T, \quad Y_i \frac{dz_i}{dx_i} + Z_i \frac{dy_i}{dx} = 0.$$

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

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

As before we have

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

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

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

We use again the bound from [11, Theorem 3.1]

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

giving

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

yielding the assertion of the Lemma. \square

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*

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

then the warmstart is successful, that is the Yildirim -Wright [17] 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

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

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

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

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 $\bar{w} \in \mathcal{N}_2(0)$, where \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

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

with

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

According to [17, Proposition 4.2] sufficient conditions for a successful warmstart are (using $\theta_0 = 0$, $\xi = \theta/2$):

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

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. \square

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

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

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.

$$(\tilde{x}_\mu^R, \tilde{y}_\mu^R, \tilde{\lambda}_\mu^R, \tilde{s}_\mu^R, \tilde{z}_\mu^R) \in \mathcal{N}_2^R(\theta) \quad (16)$$

$$(\tilde{y}_i, \tilde{\lambda}_i, \tilde{z}_i) \in \mathcal{N}_2^{(i)}(\theta) \quad (17)$$

for some $\theta \in (0, 1)$. The following two lemmas give bounds on the resulting error in the relevant components of the points compared to the exact μ -centers. Before we proceed, we need a general results stating how the components of an approximate μ -center depend on its centrality measure.

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

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

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

$$\|\tilde{x}_\mu - x_\mu\|_\infty \leq C_x(\mu) \frac{\theta}{1 - \theta},$$

$$\|\tilde{s}_\mu - s_\mu\|_\infty \leq C_s(\mu) \frac{\theta}{1 - \theta},$$

$$\|\tilde{y}_\mu - y_\mu\|_\infty \leq C_x(\mu) \frac{\theta}{1 - \theta}.$$

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

$$\begin{aligned} A^\top y(\bar{\mu}) + s(\bar{\mu}) &= c \\ Ax(\bar{\mu}) &= b \\ S(\bar{\mu})X(\bar{\mu})e &= \bar{\mu} \\ s(\bar{\mu}), x(\bar{\mu}) &> 0 \end{aligned} \quad (18)$$

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

$$\|\mu e - \tilde{\mu}\|_2 \leq \theta \mu, \quad e^\top \tilde{\mu} / n = \mu \quad (19)$$

with

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

Differentiating (18) with respect to a component $\bar{\mu}_j$ of $\bar{\mu}$ gives

$$\begin{aligned} A^\top \frac{dy(\bar{\mu})}{d\bar{\mu}_j} + \frac{ds(\bar{\mu})}{d\bar{\mu}_j} &= 0 \\ A \frac{dx(\bar{\mu})}{d\bar{\mu}_j} &= 0 \\ S(\bar{\mu}) \frac{dx(\bar{\mu})}{d\bar{\mu}_j} X(\bar{\mu}) \frac{ds(\bar{\mu})}{d\bar{\mu}_j} &= e_j, \end{aligned}$$

which we can solve for $\frac{dx(\bar{\mu})}{d\bar{\mu}_j}$, $\frac{dy(\bar{\mu})}{d\bar{\mu}_j}$, $\frac{ds(\bar{\mu})}{d\bar{\mu}_j}$ to get

$$\frac{dy(\bar{\mu})}{d\bar{\mu}_j} = -(AXS^{-1}A^\top)^{-1}AS^{-1}e_j, \quad (20a)$$

$$\frac{ds(\bar{\mu})}{d\bar{\mu}_j} = A^\top(AXS^{-1}A^\top)^{-1}AS^{-1}e_j, \quad (20b)$$

$$\frac{dx(\bar{\mu})}{d\bar{\mu}_j} = (I - S^{-1}XA^\top(AXS^{-1}A^\top)^{-1}A)S^{-1}e_j. \quad (20c)$$

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 [11, Theorem 3.1] we have the relations

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

which together with (20c) give the bound

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

For a bound on $\left\| \frac{ds(\bar{\mu})}{d\bar{\mu}_j} \right\|_\infty$ we can rewrite (20b) as

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

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

$$\begin{aligned} \left\| \frac{ds(\bar{\mu})}{d\bar{\mu}_j} \right\|_\infty &\leq \frac{1}{\mu^2(1-\theta)^2} \chi(A) \|A\|_\infty \|X\|_\infty \|S\|_\infty^2 \\ &\leq \frac{8}{\mu^2(1-\theta)^2} \chi(A) \|A\|_\infty C(d)^3 [C(d)\|d\| + \mu n]^3. \end{aligned} \quad (21)$$

Finally, for a bound on $\left\| \frac{dy(\bar{\mu})}{d\bar{\mu}_j} \right\|_\infty$ we can rewrite (20a) as

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

which yields

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

By defining

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

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

$$\left\| \frac{dx(\bar{\mu})}{d\bar{\mu}_j} \right\|_{\infty} \leq \frac{1}{1 - \theta} C_x, \quad \left\| \frac{ds(\bar{\mu})}{d\bar{\mu}_j} \right\|_{\infty} \leq \frac{1}{(1 - \theta)^2} C_s, \quad \left\| \frac{dy(\bar{\mu})}{d\bar{\mu}_j} \right\|_{\infty} \leq \frac{1}{1 - \theta} C_x.$$

Together with (19) we get

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

□

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

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

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 $(\tilde{x}_{\mu}^R, \tilde{s}_{\mu}^R)$ by

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

In the second step of the algorithm we will not find the exact μ -center for all subproblems $P_i(\tilde{x}_{\mu}^R)$, 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 Yildirim - Wright [17] 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 (16) and (17) we have that

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

and therefore

$$\begin{aligned} \|(\tilde{X}, \tilde{Y}_1, \dots, \tilde{Y}_n)(\tilde{S}, \tilde{Z}_1, \dots, \tilde{Z}_n)e - \bar{\mu}e\|_2^2 &= \|\tilde{X}\tilde{S}e - \bar{\mu}e\|_2^2 + \sum_i \|\tilde{Y}_i\tilde{Z}_ie - \bar{\mu}e\|_2^2 \\ &\leq (|\mathcal{T}| + 1)\tilde{\theta}^2\bar{\mu}^2. \end{aligned} \quad (22)$$

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 (22) 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

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

Using the bounds from Lemma 3 and Lemma 6 we have

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

Moreover, using the bounds from Lemma 2 we get

$$\begin{aligned} \|\Delta c\| &\leq \|T\| \|\mathcal{T}\| \left(C_\lambda \frac{2}{\sqrt{\mu}} \sqrt{L_Q \bar{B}} \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} + C_3 \frac{\tilde{\theta}}{1 - \tilde{\theta}} \right) \\ &\quad + \frac{2}{\sqrt{\mu^3}} \sqrt{L_Q \bar{B}^3} \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} + C_3 \frac{\tilde{\theta}}{1 - \tilde{\theta}} \\ &= \left(\|T\| \|\mathcal{T}\| C_\lambda(\mu) + \frac{\bar{B}^2}{\mu} \right) \frac{2}{\sqrt{\mu}} \sqrt{L_Q \bar{B}} \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} + (\|T\| \|\mathcal{T}\| C_\lambda(\mu) + 1) C_3 \frac{\tilde{\theta}}{1 - \tilde{\theta}} \\ &\leq C_4(\mu) \sqrt{W_1(\mathcal{T}, \mathcal{T}^R)} + C_5 \frac{\tilde{\theta}}{1 - \tilde{\theta}} \end{aligned} \quad (23)$$

where

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

Proposition 4.2 of [17] 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

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

which can be combined to obtain

$$\|\Delta c\|_\infty \leq \frac{\theta - \tilde{\theta}\sqrt{|\mathcal{T}|+1}}{2\overline{C(d)}} \min \left\{ \frac{\|d\|}{2n+1}, \frac{\mu}{4\overline{C(d)}} \right\}. \quad (24)$$

Combining (23) and (24) we get as the condition for a successful warmstart

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

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

5 Implementation and numerical results

In this section we first 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 multistage stochastic program with a given event tree \mathcal{T} , we first generate a reduced tree \mathcal{T}^R with as many scenarios as there are second-stage nodes (so that there is one scenario per branch). The deterministic equivalent problem associated to \mathcal{T}^R is solved to a specified tolerance, and the primal and dual values corresponding to the first-stage variables are stored. At this point, we instantiate a subproblem for each second-stage node by using the first-stage approximation just computed. Each subproblem is then solved serially to a specified tolerance, and the solution makes up a part of the warm-start iterate. After all subproblems have been solved, the warm-start point is complete, and is used to start the solution of the deterministic equivalent corresponding to the original tree \mathcal{T} .

We concentrate our tests on the following four problems:

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
pltxpA4-6	4	216	26,894	70,364	143,059
swing8-4	8	65,536	262,142	349,522	786,422

Table 1: Characteristics of the test problems.

The first three display stochasticity in the right-hand side, while the last one in the objective.

We tested various other combinations of tolerances (5.0e-1, 5.0e-3 and 5.0e-5 in the reduced problem, 5.0e-1, 5.0e-2 and 5.0e-3 in the subproblems). The differences among the various choices

were small and problem dependent. If the reduced tree provides a good approximation to the complete tree, then it makes sense to solve the related problem with more precise tolerance, as this will produce a better first-stage approximation.

In results shown, we use a tolerance of $5.0e-3$ for both the reduced problem and the decomposed subproblems, as this seemed to be the best combination for the tests adopted. The solutions to these are used to create the warm-start point for the complete problem, which is then solved to tolerance of $1.0e-7$. Computations were performed on a Linux PC with a 3.0GHz Intel Core 2 processor and 3GB of RAM.

In the table of results, “cold” refers the cold-started algorithm corresponding to Mehrotra’s starting point heuristic; “warm” refers to using the approach of [4] in which we choose two representative scenarios and solve the reduced problem with accuracy of $5.0e-1$; “decomp” refers to using the decomposition-based warm-start with reduced and subproblem tolerances of $5.0e-3$.

Problem	cold		warm		decomp	
fxm3-6	24	3.7	9	1.4	11	3.9
fxm3-16	62	46.2	21	18.7	19	37.0
pltxpA4-6	68	30.3	—	—	40	23.0
swing8-4	35	69.8	66	114.9	14	72.9

Table 2: Number of iterations and time in seconds to solve the complete problem to optimality for different starting points.

The results of Table 2 indicate that the warm-start scheme by decomposition is effective in the problems considered in terms of number of iterations saved compared to a cold start (in the region of 40–60%). This indicates that the scheme is capable of generating better starting points, and therefore the theoretical analysis of Section 4 is validated by a practical implementation.

In terms of computational time, compared to the warm-start scheme of [4], the decomposition approach provides major improvements only in the two largest problems, but in one case this is still not successful compared to the cold-start. We attribute this weakness of the decomposition warm-start to a shortcoming of the current implementation, which could be remedied by adopting the following observations.

First of all, the subproblems are completely independent but are currently solved serially. As we have already noted, parallelism is straightforward, and should provide close to linear speedups. This is expected to provide measurable savings on the time spent solving the subproblems without any decrease in the quality of the warm-start point generated.

Moreover, we observe that at the moment the subproblems are solved with a cold-start iterate. However, an approximate initial solution is already available from the reduced problem, and this could be used as the warm-start iterate following the approach of [4]. Again, this should yield a decrease in number of iterations needed to solve the subproblems, which may translate in time savings. We note that since the subproblems are solved with a loose tolerance, the effect of the choice of initial point is magnified, and therefore the warm-start iterate generated will not be identical to the one generated otherwise.

Finally, we currently set up a decomposed subproblem for each second-stage node. This can be relaxed by choosing only some representative second-stage nodes to spawn a subproblem. Such hybrid approach should provide a viable compromise between closeness of the warm-start iterate

to the solution and computational effort in finding it. While this would produce time savings (especially in the serial case), the quality of the warm-start point could be affected.

5.1 Quality of the starting point

One of the elements of the successful starting point is its centrality. This is defined as the distance of the complementarity pairs from the barrier parameter μ . The precise definition depends on the neighbourhood employed. The implementation of OOPS does not adopt any specific neighbourhood, but only enforces the positivity of the iterates. Therefore, we measure centrality in terms of infeasibility of the iterate with respect to the complementarity conditions $XZe = \mu e$.

In Table 3 we report the infeasibilities with respect to the various choices of starting point considered, that is

$$\xi_b = \|Ax - b\|_\infty, \quad \xi_c = \|A^\top y + z - c\|_\infty, \quad \xi_\mu = \|XZe - \mu e\|_\infty.$$

Problem		cold	warm	decomp
fxm3-6	ξ_b	2.10e+6	2.10e+2	9.00e-3
	ξ_c	6.27e+2	7.41e-4	4.28e+0
	ξ_μ	1.09e+6	1.06e+2	1.99e-2
fxm3-16	ξ_b	2.48e+6	1.90e+2	5.02e-3
	ξ_c	7.36e+2	1.16e-4	8.79e-1
	ξ_μ	1.47e+6	1.39e+1	3.78e-1
pltexpA4-6	ξ_b	1.04e+6	7.93e+2	1.12e-5
	ξ_c	5.63e+1	1.71e-5	2.01e+1
	ξ_μ	6.86e+3	2.81e-2	7.72e-4
swing8-4	ξ_b	5.56e+1	6.84e-5	3.14e-7
	ξ_c	7.21e+1	5.44e+0	5.76e-3
	ξ_μ	9.00e+3	4.84e+0	5.82e-3

Table 3: Infeasibilities of the initial point.

As expected, with the decomposition-based approach we are able to produce points that are more central than using a cold-start point or a reduced-tree warm-start scheme.

To further describe the quality of the warm-start point that we generate, we computed the distance between the initial point and the solution. We measure the infinity norms for (x, y, z) , and report the results for the different starting points. In case of cold start, the initial point is the one generated by applying Mehrotra’s starting point heuristic [9]; otherwise, we consider it to be the warm-start iterate. The column heading have the same meaning as for the previous tables. The results presented in Table 4 confirm the expectations regarding the higher quality of the warm-start point generated by the decomposition-based approach.

6 Conclusions and future research

In this paper we proposed a technique to generate a warm-start point for interior point methods that can be applied to multi-stage stochastic linear programs. The approach exploits the near-optimal solution to decomposed subproblems rooted at the second-stage nodes of the event tree

Problem		cold	warm	decomp
fxm3-6	x	13806	15266	384
	y	17726	1691	396
	z	21627	2078	2035
fxm3-16	x	30799	8395	1081
	y	12844	663	34
	z	16230	945	330
pltexpA4-6	x	1563	—	160
	y	450	—	11
	z	450568	—	11203
swing8-4	x	198	186	58
	y	20100	2480	0.01
	z	20039	6484	0.1

Table 4: Infinity norm of the distance between the initial point and the solution.

to set up an advanced iterate for the deterministic equivalent corresponding to the complete scenario tree. We observed that the iterate generated from the subproblem provides a starting point that is closer to the optimal solution, and in general provides a decrease in the number of iterations needed to reach optimality.

While at the moment this technique does not provide computational time savings, some enhancements have been suggested and can be implemented in a straightforward way.

References

- [1] J. R. BIRGE, *Decomposition and partitioning methods for multistage stochastic linear programs*, Operations Research, 33 (1985), pp. 989–1007.
- [2] J. R. BIRGE AND F. LOUVEAUX, *Introduction to stochastic programming*, Springer Series in Operations Research, New York, 1997.
- [3] J. BLOMVALL AND P. O. LINDBERG, *A Riccati-based primal interior point solver for multistage stochastic programming*, European Journal of Operational Research, 143 (2002), pp. 452–461.
- [4] M. COLOMBO, J. GONDZIO, AND A. GROTHEY, *A warm-start approach for large-scale stochastic linear programs*, Technical Report MS-06-004, School of Mathematics, The University of Edinburgh, August 2006. Accepted for publication in Mathematical Programming.
- [5] I. I. DIKIN, *On the speed of an iterative process*, Upravlaemye Sistemy, 12 (1974), pp. 54–60.
- [6] J. DUPAČOVÁ, N. GRÖWE-KUSKA, AND W. RÖMISCH, *Scenario reduction in stochastic programming*, Mathematical Programming, 95 (2003), pp. 493–511.
- [7] J. GONDZIO AND R. SARKISSIAN, *Parallel interior-point solver for structured linear programs*, Mathematical Programming, 96 (2003), pp. 561–584.
- [8] P. KALL AND S. W. WALLACE, *Stochastic programming*, John Wiley and Sons, New York, 1994.

- [9] S. MEHROTRA, *On the implementation of a primal-dual interior point method*, SIAM Journal on Optimization, 2 (1992), pp. 575–601.
- [10] J. M. MULVEY AND A. RUSZCZYŃSKI, *A new scenario decomposition method for large-scale stochastic optimization*, Operations Research, 43 (1995), pp. 477–490.
- [11] M. A. NUNEZ AND R. M. FREUND, *Condition measures and properties of the central trajectory*, Mathematical Programming, 83 (1998), pp. 1–28.
- [12] W. RÖMISCH AND R. SCHULTZ, *Stability analysis for stochastic programming*, Annals of Operations Research, 30 (1991), pp. 241–266.
- [13] R. SCHULTZ, *Some aspects of stability in stochastic programming*, Annals of Operations Research, 100 (2000), pp. 55–84.
- [14] 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.
- [15] 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.
- [16] S. J. WRIGHT, *Primal-dual interior-point methods*, SIAM, Philadelphia, 1997.
- [17] 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.