A MULTI-STEP INTERIOR POINT WARM-START APPROACH FOR LARGE-SCALE STOCHASTIC LINEAR PROGRAMMING*

MARCO COLOMBO† AND ANDREAS GROTHEY†‡

Abstract. Interior point methods (IPM) have been recognised as an efficient approach for the solution of large scale stochastic programming problems due to their ability of exploiting the block-angular structure of the augmented system particular to this problem class. Stochastic programming problems, however, have exploitable structure beyond the simple matrix shape: namely the scenarios are typically a discrete sampling of an underlying (continuous) probability distribution. An appealing way of exploiting this would be to initially use a coarser discretisation, i.e. less scenarios, to obtain an approximate solution, which could then be used to warm-start the solver on the full problem.

In this paper we present a multi-step warm-start scheme for stochastic programming problems, where a sequence of problems defined over scenario trees of differing sizes is given and an IPM warm-start point is constructed by successively finding approximations to the central path of the problems defined over the given trees. We analyse the resulting algorithm, argue that it yields improved complexity over either the coldstart or a naive two-step scheme, and give numerical results.

Keywords: Stochastic Programming, Interior-point Methods, Warm-starting.
Classification: 90C51, 90C15.

1. Introduction. A stochastic programming problem is defined on a tree $\mathcal{T}$. This *target* tree encodes the available information about the future as well as the decision structure of the problem. We intend to approximate the problem by considering a sequence of trees $\mathcal{T}^{(k)}$ with $\mathcal{T}^{(k)} \to \mathcal{T}$ in some specified sense. We use the notation of $P(\mathcal{T})$ for the stochastic program on a given tree $\mathcal{T}$. The guiding idea will be to use some information obtained by solving $P(\mathcal{T}^{(k)})$ to speed up the solution for $P(\mathcal{T}^{(k+1)})$, eventually leading to a faster solution of the target problem $P(\mathcal{T})$. Since we are using interior point methods to solve the problems, we will not attempt to find the optimal solution for each intermediate problem. Rather, for every tree in the sequence we find an approximate solution $(x^{(k)}, y^{(k)}, s^{(k)}) \in \mathcal{N}_*^{(k)}$, where $\mathcal{N}_*$ is an appropriately chosen neighbourhood of the central path. We intend to do this in such a way that the approximate solution for tree $\mathcal{T}^{(k)}$ can be used to warm-start an interior point method on tree $\mathcal{T}^{(k+1)}$, possibly after performing a modification step, which might involve inter-/extrapolation of scenarios, a more traditional modification step in the sense of [10, 4], or a combination thereof.

In an earlier paper [3] the authors suggested the use of one smaller (reduced) tree $\mathcal{T}_R$ to speed up the solution on the target tree. The contribution of this paper is to extend this idea to a sequence of trees, as well as providing some detailed analysis of the method and arguing for the improved complexity of the multi-tree scheme over either a coldstart or a two-step scheme. It turns out, unfortunately, that the analysis presented in [3] for the two-step scheme is not directly applicable to the multi-tree situation. This is for two reasons: first in the multi-tree scheme problem parameters such as the size, condition number and the value of the barrier parameter μ at which the warmstart is performed change from step to step, and the analysis in [3] does not make the dependence of the conditions for a successful warmstart on these parameters explicit. Secondly the earlier paper imposes conditions on the tree-

*This research has been supported by EPSRC grant EP/E036910/1.

†School of Mathematics and Maxwell Institute, The University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK.
‡A.Grothey@ed.ac.uk

balancing parameter ρ which are too tight to be useful in a multi-step scheme. It turns out that these tight conditions are based on the use of the $\mathcal{N}_s$-neighbourhood, and can be substantially relaxed in the $\mathcal{N}_{-\infty}$ setting.

The main contribution of this paper over [3] is thus to generalise results for the two-step scheme in a way that are appropriate for the multi-tree scheme, to propose how to manage such a multi-tree scheme and to demonstrate theoretical and practical efficiency gains of such a scheme over the two-step warmstart.

In the following section we review the basic background of stochastic programming and IPM, mainly to introduce our notation. In Section 3 we review some warm-start results for IPM and adapt them to our setting. Section 4 analyses a single step of the multi-tree scheme, whereas in Section 5 we give the main results for the multi-tree warm-start scheme. Finally in Section 6 we present numerical results.

2. Notation. A (linear, two-stage) stochastic programming problem is a mathematical programming problem in which some of the data is unknown a-priori:

$$\begin{array}{rl} \min & c^T x + I\!E_\xi[Q(x,\xi)] \\ \text{s.t.} & Ax = b, \quad x \geq 0 \\ \text{where} & Q(x,\xi) = \min\{q(\xi)^T y(\xi) : W(\xi)y(\xi) = h(\xi) - T(\xi)x, y(\xi) \geq 0\}. \end{array} \tag{2.1}$$

Here ξ is the random variable representing the uncertainty in the data. $Q(x,\xi)$ is the recourse function depending on the *first stage decisions* x and the random event ξ, whereas $y(\xi)$ are the second stage (or recourse) decisions. Note that the stochastic programming model implies a partition of the decision variables (x,y) into a first stage x which have to be independent of the observed event ξ and a second stage y whose values can depend on the observed value of ξ. Problem (2.1) can be rewritten [8, Thm 2.20]

$$\begin{array}{rl} \min & c^T x + I\!E_\xi[q(\xi)^T y(\xi)] \\ \text{s.t.} & Ax = b, \\ & W(\xi)y(\xi) = h(\xi) - T(\xi)x, \quad \text{a.e.} \\ & x \geq 0, \quad y(\xi) \geq 0. \end{array} \tag{2.2}$$

It is usually assumed that the distribution ξ affecting the random data is known. In general ξ is a continuous distribution, in which case solving (2.2) becomes very challenging. To make it computationally tractable, ξ is replaced by a discrete approximation $\tilde{\xi}$. If $\tilde{\xi}$ takes values $\tilde{\xi}_i$ with probabilities $P(\tilde{\xi} = \tilde{\xi}_i) = p_i$, $i \in \mathcal{I}$, then (2.2) can be rewritten as the *deterministic equivalent*

$$\begin{array}{rl} \min & c^T x + \sum_i p_i q_i^T y_i \\ \text{s.t.} & Ax = b, \\ & T_i x + W_i y_i = h_i, \quad i \in \mathcal{I}, \\ & x \geq 0, y_i \geq 0, \end{array} \tag{2.3}$$

where $q_i = q(\tilde{\xi}_i)$ and analogously for T_i, W_i, h_i, y_i. We denote by $\eta_i = (T_i, W_i, h_i, q_i)$ the data describing the i-th scenario and use a notional 0-th scenario $\eta_0 = (A, b, c)$ to denote the data corresponding to the first stage. By analogy with multistage stochastic programming we will borrow the term *tree* to denote a scenario set

$$\mathcal{T} := \{\eta_i : i \in \mathcal{I}\} \cup \{\eta_0\}.$$

Since the tree $\mathcal{T}$ contains all information to describe the stochastic programming problem we will use the notation $P(\mathcal{T})$ for problem (2.3) on tree $\mathcal{T}$. In this paper

we assume that we are given a series of successively larger trees $\mathcal{T}^{(1)}, \ldots, \mathcal{T}^{(K)} = \mathcal{T}$, each of which serves as an approximation to the next tree in the sequence. To analyse how well a given tree $\mathcal{T}^{(k+1)}$ is approximated by the previous tree $\mathcal{T}^{(k)}$ we need to define the concept of a distance between two trees. As in [3], we base this on *scenario distances*, which we define as follows:

DEFINITION 2.1. *The* **distance between two scenarios** $\eta_i, \eta_j \in \mathcal{T} \setminus \{\eta_0\}$ *is*

$$d(\eta_i, \eta_j) := \|T_i - T_j\|_2 + \|W_i - W_j\|_2 + \|h_i - h_j\|_2 + \|q_i - q_j\|_2.$$

The **distance of a scenario** η **from a tree** $\mathcal{T}$ *is thus*

$$d(\eta, \mathcal{T}) := \min_{\eta_i \in \mathcal{T}} d(\eta, \eta_i).$$

Therefore, the **distance between two trees** $\mathcal{T}_i$ *and* $\mathcal{T}_j$ *can be defined as*

$$d(\mathcal{T}_i, \mathcal{T}_j) := \max_{\eta_i \in \mathcal{T}_i} d(\eta_i, \mathcal{T}_j). \tag{2.4}$$

2.1. Interior Point Methods. Interior point methods have become a standard method for the solution of large scale stochastic programming problems. Their performance is similar to dedicated active-set based approaches such as the L-shaped method but, unlike these, they are also applicable to nonlinear formulations. IPMs applied to solve the linear problem

$$\min\ c^T x \quad \text{s.t.}\ \ Ax = b,\ x \geq 0 \tag{2.5}$$

proceed by replacing the $x \geq 0$ constraints in (2.5) by logarithmic barrier terms added to the objective function to arrive at

$$\min\ c^T x - \mu \sum_i \ln x_i \quad \text{s.t.}\ \ Ax = b. \tag{2.6}$$

We define the set of strictly primal–dual feasible points for (2.6) as

$$\mathcal{F}^0 := \{(x, y, s) : Ax = b, A^T y + s = c, (x, s) > 0\}.$$

Assuming $\mathcal{F}^0 \neq \emptyset$, then for every $\mu > 0$ problem (2.6) is a strictly convex optimization problem, whose unique solution is given by the solution to the KKT-equation system

$$\begin{array}{rcl} Ax & = & b \\ A^T y + s & = & c \\ XSe & = & \mu e \\ x, s & > & 0, \end{array} \tag{2.7}$$

where we used the notation $X = \text{diag}(x_1, ..., x_n)$, $S = \text{diag}(s_1, ..., s_n)$, $e = (1, ..., 1)^T$. The solutions of (2.7) for different values of $\mu > 0$ trace a trajectory (the *central path*) that reaches the solution of (2.5) for $\mu = 0$. In the analysis of interior point methods it is common to work with a neighbourhood of the central path. We define the following:

$$\begin{aligned} \mathcal{N}_2(\theta) &:= \{(x, y, s) \in \mathcal{F}^0 | \|XSe - \mu e\|_2 \leq \theta\mu, \mu := x^T s/n\} \\ \mathcal{N}_{-\infty}(\gamma) &:= \{(x, y, s) \in \mathcal{F}^0 | \gamma\mu \leq x_i s_i, \mu := x^T s/n\} \\ \mathcal{N}_s(\gamma) &:= \{(x, y, s) \in \mathcal{F}^0 | \gamma\mu \leq x_i s_i \leq \mu/\gamma, \mu := x^T s/n\} \end{aligned}$$

Of these $\mathcal{N}_2$ and $\mathcal{N}_{-\infty}$ are in common use, whereas $\mathcal{N}_s$ is used in [3, 4].

Interior point methods can be seen as homotopy methods that alternate between steps of Newton's method to solve the nonlinear system (2.7) and reductions of the barrier parameter μ towards zero. The Newton steps to solve (2.7) are of the form

$$A\Delta x = \xi_b := b - Ax \tag{2.8a}$$
$$A^T\Delta y + \Delta s = \xi_c := c - A^T y - s \tag{2.8b}$$
$$X\Delta s + S\Delta x = r_{xs} := \mu^+ e - XSe, \tag{2.8c}$$

where μ^+ is the next target μ-value in the homotopy sequence. Appropriate globalisation strategies, such as restricting steps to be within a neighbourhood of the central path, are usually employed.

3. Warmstarting Interior Point Methods. The warmstarting of interior point methods is generally seen to be difficult to impossible. While it is true that interior point methods are unlikely to match the warm-start efficiency of active set type methods (such as the simplex method), recent results [2, 3, 4] suggest that typically between 50%-60% of iterations can be saved by using a warm-start procedure.

By warm-starting we aim to use information gained from the solution of the LP problem (2.5), the *original problem*, to significantly speed up the solution of a *modified problem*:

$$\min\ \bar{c}^T x \quad \text{s.t.}\ \ \bar{A}x = \bar{b},\ x \geq 0, \tag{3.1}$$

which is assumed to be *close* to (2.5), that is the perturbation of problem data

$$\Delta d := (\Delta A, \Delta b, \Delta c) = (\bar{A} - A, \bar{b} - b, \bar{c} - c)$$

is small. Unlike the situation with active set based strategies, it is not a good idea to use the solution of the original problem as a starting point for the modified problem. Instead, a popular strategy is to select an advanced μ-centre (that is an approximation to a point of the central path) of problem (2.5) and to reason that (under suitable conditions, and possibly after some modifications) the obtained point is close to the central path of problem (3.1) and can hence be used successfully as an advanced iterate for this problem. This is the approach taken by [4, 5, 10]; however see [2] for a different view.

Several authors [4, 10] have looked at conditions that guarantee a successful warm-start. These conditions usually impose bounds on the size of Δd or equivalently the implied distance from feasibility of the warm-start point in the modified problem, the centrality of the warm-start point and the size of μ as a measure of distance to optimality. Typically larger changes in problem data require a warm-start point corresponding to a larger value of μ.

In the rest of this section we will give variants of the warm-start results in [4] that are adapted to our setting. Later we will use these to analyse the complete multi-tree scheme.

3.1. Results based on implied infeasibilities. The setup considered in [4] takes an advanced iterate (x, y, s) encountered during the solution of (2.5) as the basis for the warm-start point. Since the problem data has changed, (x, y, s) is not primal–dual feasible in the modified problem (3.1) but instead implies residuals

$$\bar{\xi}_b = \bar{b} - \bar{A}x, \qquad \bar{\xi}_c = \bar{c} - \bar{A}^T y - s.$$

Here and in the rest of the section we use the notation that *bared* entities (such as $\bar{\xi}, \bar{\mathcal{N}}_2$) are associated with the modified problem. To re-gain feasibility a *modification step* is added, which is based on solving the Newton system (2.8) with $r_{xs} = 0$ (i.e. the modification step attempts to gain feasibility but does not seek to improve centrality). The analysis in [4] considers under what conditions the full modification step can be taken and hence absorbs the residuals, leading to a primal–dual warm-start point in the modified problem. The results are based on bounds on the implied residuals

$$\delta_{bc} := \|\bar{\xi}_c\|_2 + \|\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{\xi}_b\|_2, \tag{3.2}$$

or, rather, its scaled version

$$\tilde{\delta}_{bc} := \|S^{-1}\bar{\xi}_c\|_2 + \|X^{-1}\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{\xi}_b\|_2 = \|\tilde{\xi}_c\|_2 + \|\tilde{\xi}_b\|_2.$$

However, the analysis in [4] assumes that the perturbations in problem data will only affect *either* $\bar{\xi}_b$ *or* $\bar{\xi}_c$ but not both at the same time. Moreover, the analysis only gives conditions under which a full modification step is feasible (i.e. a primal–dual feasible point for the modified problem is obtained), but makes no claims about the centrality of this point. In [3, Thm 6] on the other hand, such an analysis is given, but only for the $\mathcal{N}_s$-neighbourhood and based on the scaled residuals $\tilde{\delta}_{bc}$, which hides the μ-dependence of the resulting bounds. The aim of this section is to generalise the results in a way that addresses these issues.

REMARK 1. *The term $\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{\xi}_b$ that appears in the definition of δ_{bc} is the "orthogonal distance of x from primal feasibility ($\{x : \bar{A}x = \bar{b}\}$)":*

$$\begin{aligned}\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{\xi}_b &= \bar{A}^T(\bar{A}\bar{A}^T)^{-1}(\bar{b} - \bar{A}x)\\ &= [(I - \bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{A})x + \bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{b}] - x\\ &= P_{\{x:\bar{A}x=\bar{b}\}}(x) - x,\end{aligned}$$

where $P_{\{x:\bar{A}x=\bar{b}\}}(x)$ is the orthogonal projection of x onto the subspace $\{x : \bar{A}x = \bar{b}\}$. Similarly, we can interpret

$$\bar{\xi}_c = \bar{c} - \bar{A}^T y - s = P_{\{s:\exists y:\bar{A}^T y+s=\bar{c}\}}(s) - s$$

as the orthogonal distance of s from dual feasibility. In other words δ_{bc} measures the total orthogonal distance of the warm-start point (x, s) from primal–dual feasibility in the modified problem.

The results in [4] are based on the observation that, using (2.8), the modification step satisfies

$$\Delta x = -XQ\tilde{\xi}_c + X(I - Q)\tilde{\xi}_b \qquad \text{and} \qquad \Delta s = SQ\tilde{\xi}_c - S(I - Q)\tilde{\xi}_b,$$

that is

$$X^{-1}\Delta x = -Q\tilde{\xi}_c + (I - Q)\tilde{\xi}_b = -S^{-1}\Delta s, \tag{3.3}$$

where

$$\begin{aligned}Q &= I - S^{-1}\bar{A}^T(\bar{A}XS^{-1}\bar{A}^T)^{-1}\bar{A}X,\\ \tilde{\xi}_c &= S^{-1}\bar{\xi}_c = S^{-1}(\bar{c} - \bar{A}^T y + s),\\ \tilde{\xi}_b &= X^{-1}\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{\xi}_b = X^{-1}\bar{A}^T(\bar{A}\bar{A}^T)^{-1}(\bar{b} - \bar{A}x).\end{aligned}$$

As in [4], we can derive bounds on $\|Q\|_2$ that are used in conditions on a successful warm-start.

LEMMA 3.1. *We have the following bounds for* $\|Q\|_2$*:*

$$(x,y,s) \in \mathcal{N}_2(\theta), \quad \Rightarrow \|Q\|_2 \leq \left(\frac{1+\theta}{1-\theta}\right)^{1/2}$$
$$(x,y,s) \in \mathcal{N}_{-\infty}(\gamma), \quad \Rightarrow \|Q\|_2 \leq 1 + (\frac{1-\gamma}{\gamma})^{1/2}\sqrt{n}.$$

Also, if a bound $B_\infty : x_i, s_i < B_\infty$ *is known, we have*

$$(x,y,s) \in \mathcal{N}_{-\infty}(\gamma), \quad \Rightarrow \|Q\|_2 \leq \frac{B_\infty}{\sqrt{\gamma\mu}}.$$

Proof. The first result is from [4, Lemma 3.1/3.2]. For the other results note that Q can be written as

$$Q = X^{-1/2}S^{-1/2}PX^{1/2}S^{1/2}, \quad P = I - X^{1/2}S^{-1/2}\bar{A}^T(\bar{A}XS^{-1}\bar{A}^T)^{-1}\bar{A}X^{1/2}S^{-1/2} \tag{3.4}$$

with P representing the orthogonal projection onto the null space of $\bar{A}X^{1/2}S^{-1/2}$ and hence $\|P\|_2 = 1$. Therefore

$$\|Q\|_2 \leq \|X^{-1/2}S^{-1/2}\|_2\|X^{1/2}S^{1/2}\|_2.$$

For $(x,y,s) \in \mathcal{N}_{-\infty}(\gamma)$ we have $x_i s_i \geq \gamma\mu$ and $x_i s_i = \sum_j x_j s_j - \sum_{j\neq i} x_j s_j \leq n\mu - (n-1)\gamma\mu \leq (1-\gamma)n\mu + \gamma\mu$ and hence

$$\|Q\|_2 \leq 1 + (\frac{1-\gamma}{\gamma})^{1/2}\sqrt{n}.$$

Alternatively, in presence of the bound $x_i, s_i < B_\infty$ we have $\|X_i^{1/2}S_i^{1/2}\|_2 \leq B_\infty$, so $\|Q\|_2 \leq \frac{B_\infty}{\sqrt{\gamma\mu}}$. ∎

We now give a generalisation of the warm-start results in [3, 4] that applies in the presence of both primal and dual residuals and is not tied to a particular neighbourhood.

LEMMA 3.2. *Let* (x,y,s) *be the considered warm-start point satisfying* $x_i, s_i \leq B_\infty, x_i s_i \geq \gamma\mu, \forall i$*, where* $\mu = x^T s/n$*. If*

$$\delta_{bc} \leq \frac{\mu\gamma}{B_\infty(1+\|Q\|_2)},$$

then the full modification step from the warm-start iterate can be taken and absorbs the complete primal–dual infeasibilities.

Proof. Sufficient for a successful warm-start is $x + \Delta x \geq 0, s + \Delta s \geq 0$, that is

$$\|X^{-1}\Delta x\|_\infty \leq 1, \quad \|S^{-1}\Delta s\|_\infty \leq 1.$$

From (3.3) we have

$$\|X^{-1}\Delta x\|_2 \;\leq\; \|Q\|_2\|\tilde{\xi}_c\|_2 + (1+\|Q\|_2)\|\tilde{\xi}_b\|_2 \;\leq\; (1+\|Q\|_2)(\|\tilde{\xi}_c\|_2 + \|\tilde{\xi}_b\|_2).$$

Under the condition of the Lemma we get

$$\|\tilde{\xi}_c\|_2 = \|S^{-1}\bar{\xi}_c\|_2 \leq \frac{1}{\min_i s_i}\|\bar{\xi}_c\|_2 \leq \frac{B_\infty}{\mu\gamma}\|\bar{\xi}_c\|_2,$$

since $x_i s_i \geq \gamma\mu$, $\|x\|_\infty \leq B_\infty$ and therefore $s_i \geq \gamma\mu/x_i \geq \gamma\mu/B_\infty$. Similarly, we have

$$\|\tilde{\xi}_b\|_2 = \|X^{-1}\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{\xi}_b\|_2 \leq \frac{1}{\min_i x_i}\|\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{\xi}_b\|_2 \leq \frac{B_\infty}{\mu\gamma}\|\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{\xi}_b\|_2.$$

Considering the two terms together we obtain

$$\|S^{-1}\Delta s\|_2 = \|X^{-1}\Delta x\|_2 \leq (1 + \|Q\|_2)\frac{B_\infty}{\mu\gamma}(\|\bar{\xi}_c\|_2 + \|\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{\xi}_b\|_2),$$

so that $\|S^{-1}\Delta s\| = \|X^{-1}\Delta x\| \leq 1$ provided that

$$\delta_{bc} = \|\bar{\xi}_c\|_2 + \|\bar{A}^T(\bar{A}\bar{A}^T)^{-1}\bar{\xi}_b\|_2 \leq \frac{\mu\gamma}{B_\infty(1 + \|Q\|_2)}.$$

□

We are now ready to give two extensions of this result that also make statements about the centrality of the resulting point. These should be compared with [3, Thm 6] for the $\mathcal{N}_s$-neighbourhood.

LEMMA 3.3. *Let $(x, y, s) \in \mathcal{N}_2(\theta_0)$ be the warm-start iterate. Let $\theta : 0 < \theta_0 < \theta < 1$ be given and $n > 9$. Let the modification step $(\Delta x, \Delta y, \Delta s)$ be obtained by solving system (2.8) with $r_{xs} = 0$ in the new problem for the current value of μ. If we have*

$$\delta_{bc} \leq \frac{\beta\mu(1 - \theta_0)}{B_\infty(1 + \|Q\|_2)}, \tag{3.5}$$

with

$$\beta \leq \sqrt{\frac{\theta - \theta_0}{3\sqrt{n}(1 + \theta_0)}}, \tag{3.6}$$

then the full modification step is feasible, absorbs all infeasibilities and satisfies

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

Proof. As before we need to prove that the complementarity products $\bar{x}_i\bar{s}_i$ do not deviate too much from the average. Note that under the conditions of the Lemma we have from a slight generalisation of Lemma 3.2 that

$$\|X^{-1}\Delta x\|_\infty \leq \beta, \qquad \|S^{-1}\Delta s\|_\infty \leq \beta,$$

and therefore

$$\Delta x_i \leq \beta x_i, \quad \Delta s_i \leq \beta s_i. \tag{3.7}$$

We also rely on the fact that for $(x, y, s) \in \mathcal{N}_2(\theta_0)$ we have

$$|x_i s_i - \mu| \leq \theta_0\mu, \qquad x_i s_i \leq (1 + \theta_0)\mu,$$

and from (2.8c)

$$x_i\Delta s_i + s_i\Delta x_i = 0$$

The rest of the proof follows that of [10, Proposition 4.2]. We start by finding a bound on the norm of the vector

$$[(x_i + \Delta x_i)(s_i + \Delta s_i)]_{i=1,2,\ldots,n} - [(x + \Delta x)^T (s + \Delta s)/n]e. \tag{3.8}$$

Using the above relations we can bound $\Delta x_i \Delta s_i = x_i s_i \Delta s_i / s_i \Delta x_i / x_i \leq \beta^2(1 + \theta_0)\mu$ and thus

$$\|\Delta X \Delta S e\|_2 \leq \sqrt{n}\beta^2(1 + \theta_0)\mu, \qquad \|[\Delta x^T \Delta s/n]e\|_2 \leq \sqrt{n}\beta^2(1 + \theta_0)\mu$$

which can be used to produce a bound on (3.8) by

$$\begin{aligned}
\|(X+ \ \Delta X)(S + \Delta S)e - \bar{\mu}e\|_2 & \\
&= \|(X + \Delta X)(S + \Delta S)e - [(x + \Delta x)^T (s + \Delta s)/n]e\|_2 \\
&\leq \|XSe - \mu e\|_2 + \|X\Delta S e + S\Delta X e\|_2 + \|\Delta X \Delta S e\|_2 \\
&\qquad + \|[x^T \Delta s + s^T \Delta x/n]e\|_2 + \|[\Delta x^T \Delta s/n]e\|_2 \\
&\leq \theta_0 \mu + 2\sqrt{n}\beta^2(1 + \theta_0)\mu.
\end{aligned}$$

Meanwhile, we obtain a lower bound on the duality measure after the correction by

$$\bar{\mu} = (x + \Delta x)^T (s + \Delta s)/n \geq \mu - \beta^2(1 + \theta_0)\mu.$$

Therefore, a sufficient condition for $(x + \Delta x, y + \Delta y, s + \Delta s) \in \bar{\mathcal{N}}_2(\theta)$ is that

$$\theta_0 \mu + 2\sqrt{n}\beta^2(1 + \theta_0)\mu \leq \theta\mu - \beta^2\theta(1 + \theta_0)\mu,$$

which after rearrangement becomes

$$\mu(\theta - \theta_0) \geq 2\sqrt{n}\beta^2(1 + \theta_0)\mu + \beta^2\theta(1 + \theta_0)\mu = \beta^2(2\sqrt{n} + \theta)(1 + \theta_0)\mu. \tag{3.9}$$

For n large enough, $\sqrt{n} > \theta$ and therefore (3.6) is sufficient for (3.9). □

COROLLARY 3.4. *Under the setup of Lemma 3.3 we have* $(\bar{x}, \bar{y}, \bar{s}) \in \bar{\mathcal{N}}_2(\theta)$ *if*

$$\delta_{bc} \leq \frac{(1 - \theta_0)^{3/2}(\theta - \theta_0)^{1/2}}{8\sqrt[4]{n}B_\infty}\mu.$$

Proof. From Lemma 3.1 we have

$$\|Q\|_2 \leq \left(\frac{1 + \theta_0}{1 - \theta_0}\right)^{1/2}$$

which, combined with (3.5) and (3.6), yields the sufficient condition

$$\delta_{bc} \leq \frac{(\theta - \theta_0)^{1/2}}{\sqrt{3}\sqrt[4]{n}B_\infty(1 + \theta_0)^{1/2}} \frac{(1 - \theta_0)}{\left(1 + (\frac{1+\theta_0}{1-\theta_0})^{1/2}\right)}\mu. \tag{3.10}$$

We have

$$\frac{1 - \theta_0}{(1 + \theta_0)^{1/2}\left(1 + (\frac{1+\theta_0}{1-\theta_0})^{1/2}\right)} = \frac{(1 - \theta_0)^{3/2}}{(1 + \theta_0)^{1/2}[(1 - \theta_0)^{1/2} + (1 + \theta_0)^{1/2}]} \geq \frac{1}{4}(1 - \theta_0)^{3/2},$$

which when combined with (3.10) proves the Corollary. □

Finally we give an equivalent of Lemma 3.3 involving the $\mathcal{N}_{-\infty}$ neighbourhood.

LEMMA 3.5. *Let $(x, y, s) \in \mathcal{N}_{-\infty}(\gamma_0)$ be the warm-start iterate. Let $\gamma : 0 < \gamma < \gamma_0 < 1$ be given. Let the modification step $(\Delta x, \Delta y, \Delta s)$ be obtained by solving system (2.8) for $r_{xs} = 0$ in the new problem for the current value of μ. If we have*

$$\delta_{bc} \leq \frac{\beta\mu\gamma_0}{B_\infty(1 + \|Q\|_2)}, \tag{3.11}$$

with

$$\beta \leq \sqrt{\frac{\gamma_0 - \gamma}{\gamma_0}}, \tag{3.12}$$

then the full modification step is feasible, absorbs all infeasibilities and satisfies $(\bar{x}, \bar{y}, \bar{s}) \in \bar{\mathcal{N}}_{-\infty}(\gamma)$.

Proof. First we note that (3.12) implies $0 < \beta < 1$ and

$$1 - \beta^2 \geq 1 - \frac{\gamma_0 - \gamma}{\gamma_0} = \gamma/\gamma_0 \tag{3.13}$$

The situation is similar to that of [3, Theorem 6] (where the $\mathcal{N}_s$- rather than the $\mathcal{N}_{-\infty}$-neighbourhood is considered. However the proof carries over to the given setup and thus we have

$$(\bar{x}, \bar{y}, \bar{s}) \in \mathcal{N}_{-\infty}((1 - \beta^2)\gamma_0) \subseteq \mathcal{N}_{-\infty}(\gamma)$$

□

COROLLARY 3.6. *Under the setup of Lemma 3.5 we have $(\bar{x}, \bar{y}, \bar{s}) \in \bar{\mathcal{N}}_{-\infty}(\gamma_0)$ if*

$$\delta_{bc} \leq \frac{\gamma_0\sqrt{\gamma_0 - \gamma}}{2B_\infty^2}\mu^{3/2}. \tag{3.14}$$

Proof. From Lemma 3.1 we have $\|Q\|_2 \leq B_\infty/(\sqrt{\gamma_0\mu})$. Also, assuming $\mu < B_\infty^2$ we have $(1 + \frac{B_\infty}{\sqrt{\gamma_0\mu}}) \leq 2\frac{B_\infty}{\sqrt{\gamma_0\mu}}$ and therefore

$$\frac{\gamma_0\sqrt{\gamma_0 - \gamma}}{2B_\infty^2}\mu^{3/2} \leq \frac{\sqrt{(\gamma_0 - \gamma)\gamma_0}}{B_\infty(1 + \frac{B_\infty}{\sqrt{\gamma_0\mu}})}\mu \leq \frac{\sqrt{(\gamma_0 - \gamma)\gamma_0}}{B_\infty(1 + \|Q\|_2)}\mu,$$

that is condition (3.14) is sufficient for (3.11). □

4. Warm-starting stochastic programming problems. We now turn our attention to the warm-starting of stochastic programming problems. In what follows we will limit our analysis to the case of a two stage stochastic programming problem with fixed recourse where the stochasticity only affects the right hand side and the recourse cost q, i.e problem (2.2) with $W(\xi) = W, T(\xi) = T$. Further we assume full recourse and that W has maximum rank.

As indicated earlier, we assume that we are given a sequence of trees $\mathcal{T}^{(1)}, ..., \mathcal{T}^{(K)}$ which define a series of problems $P^{(k)} = P(\mathcal{T}^{(k)})$. In this section we will start by analysing a single step of this scheme, i.e. we assume that we only have two trees and postpone the analysis of the whole sequence to Section 5.

This warm-start application differs from the usual setup in that, rather than warm-starting from a problem with the same size but different problem data, we are

warm-starting from a problem with the same *structure*, but different dimensions. As in [3] we will introduce a notional *expanded problem* that has the same size as the problem to be warm-started (the new problem) but using data from the old problem. As in [3] it is possible to construct a primal-dual feasible point for the expanded problem from an iterate of the reduced problem; this however results in a worsening of centrality of the point.

We now assume that we are given two trees $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}$, where $\mathcal{T}^{(1)}$ is obtained by an aggregation of nodes from $\mathcal{T}^{(2)}$. In detail, we partition the scenario set of $\mathcal{T}^{(2)}$ into disjoint clusters

$$\mathcal{T}^{(2)} = \bigcup_j C_j, \qquad C_j \cap C_i = \emptyset, \qquad \forall j \neq i$$

find the *average scenario* for each cluster

$$\mu(C_j) = \sum_{i \in C_j} p_i \eta_i / (\sum_{i \in C_j} p_i)$$

and use the set of aggregated scenarios as the reduced tree

$$\mathcal{T}^{(1)} := \{\mu(C_j)\}_j.$$

This implies a function

$$r^{(1,2)} : \mathcal{T}^{(2)} \to \mathcal{T}^{(1)}, \quad r^{(1,2)}(\eta) = \mu(C_j) : \eta \in C_j$$

that identifies for each scenario of $\mathcal{T}^{(2)}$ the corresponding scenario in $\mathcal{T}^{(1)}$, and its (set valued) inverse

$$\mathcal{I}^{(1,2)}(\nu) := \{\eta \in \mathcal{T}^{(2)} : r(\eta) = \nu\}, \quad \nu \in \mathcal{T}^{(1)}.$$

For the probabilities p_ν of scenario ν we have

$$p_\nu = \sum_{\eta \in \mathcal{I}^{(1,2)}(\nu)} p_\nu, \qquad \forall \nu \in \mathcal{T}^{(1)},$$

that is tree $\mathcal{T}^{(1)}$ is obtained from $\mathcal{T}^{(2)}$ by clustering and node aggregation. For a balanced aggregation we would have $p_\eta |\mathcal{T}^{(2)}| \approx p_{r(\eta)} |\mathcal{T}^{(1)}|$, where $|\mathcal{T}|$ is used to denote the cardinality of the tree $\mathcal{T}$. We define the balancing parameter $\rho^{(1,2)}$ to be the maximal deviation from this

$$\rho^{(1,2)} := \min_{\eta \in \mathcal{T}^{(2)}} \left\{ \frac{p_\eta}{p_{r(\eta)}} \frac{|\mathcal{T}^{(2)}|}{|\mathcal{T}^{(1)}|}, \frac{p_{r(\eta)}}{p_\eta} \frac{|\mathcal{T}^{(1)}|}{|\mathcal{T}^{(2)}|} \right\} \leq 1. \tag{4.1}$$

In addition we also need a one-sided version

$$\rho_-^{(1,2)} := \min_{\eta \in \mathcal{T}^{(2)}} \left\{ \frac{p_\eta}{p_{r(\eta)}} \frac{|\mathcal{T}^{(2)}|}{|\mathcal{T}^{(1)}|} \right\} \leq 1. \tag{4.2}$$

An important difference between $\rho^{(1,2)}$ and $\rho_-^{(1,2)}$ concerns the root node (labelled '0'). Since $r(0) = 0$ and $p_0 = 1$ for all aggregations, we have

$$\rho^{(1,2)} \leq |\mathcal{T}^{(1)}| / |\mathcal{T}^{(2)}|$$

irrespective of the balancing of the tree aggregation. On the other hand, it is possible to achieve $\rho_-^{(1,2)} = 1$ for a perfectly balanced aggregation. It turns out that the worsening of centrality is governed by ρ when using the $\mathcal{N}_s$-neighbourhood, but by ρ_- for $\mathcal{N}_{-\infty}$. This is the main reason why we chose to repeat much of the analysis of [3] for the case of $\mathcal{N}_{-\infty}$.

As in [3] we will use a two-step procedure to construct a warm-start point for problem $P(\mathcal{T}^{(2)})$ from a (reasonably) central and primal–dual feasible point for $P(\mathcal{T}^{(1)})$. In a first step, with the mapping $r^{(1,2)}$ we define the *expanded tree*

$$\hat{\mathcal{T}}^{(1,2)} := \{r^{(1,2)}(t) : t \in \mathcal{T}^{(2)}\},$$

that is a tree of the same shape as $\mathcal{T}^{(2)}$, but with all scenarios replaced by their corresponding scenarios from $\mathcal{T}^{(1)}$. When it is clear which two trees are involved we will drop the superscript $^{(1,2)}$ from r and $\hat{\mathcal{T}}$. The warm-start point can be similarly expanded resulting in a point that is primal–dual feasible for $P(\hat{\mathcal{T}})$ but has worsened centrality. From a primal–dual feasible point (x, y, s) for tree $\mathcal{T}^{(1)}$, then as in [3] we can construct a primal–dual feasible point $(\hat{x}, \hat{y}, \hat{s})$ for problem $P(\hat{\mathcal{T}})$ by setting

$$\hat{x}^t = x^{r(t)}, \qquad (\hat{y}^t, \hat{s}^t) = \frac{p_t}{p_{r(t)}}(y^{r(t)}, s^{r(t)}). \tag{4.3}$$

In a second step we replace the scenarios of $\hat{\mathcal{T}}$ by those of $\mathcal{T}^{(2)}$ resulting in implied residuals ξ_b, ξ_c for the constructed warm-start point. We can control the worsening of centrality (in step 1) by using a balanced aggregation, and the size of the implied residuals (in step 2) by keeping $d(\mathcal{T}^{(1)}, \mathcal{T}^{(2)})$ small. The main result of this section are conditions on $\rho^{(1,2)}$ and $d(\mathcal{T}^{(1)}, \mathcal{T}^{(2)})$ that guarantee a successful warm-start.

4.1. Step 1: Worsening of centrality. We start by analysing the worsening of centrality. The corresponding result from [3] (where the $\mathcal{N}_s$-neighbourhood is used), is the following.

LEMMA 4.1 ([3, Theorem 3]). *Let $(x, y, s) \in \mathcal{N}_s^{(1)}(\gamma)$ and let $(\hat{x}, \hat{y}, \hat{s})$ be constructed from it by (4.3). Then*

$$(\hat{x}, \hat{y}, \hat{s}) \in \hat{\mathcal{N}}_s(\rho^{(1,2)}\gamma).$$

We now give similar results for the $\mathcal{N}_2$ and $\mathcal{N}_{-\infty}$ neighbourhoods.

LEMMA 4.2. *Let $(x, y, s) \in \mathcal{N}_2^{(1)}(\theta)$ and let $(\hat{x}, \hat{y}, \hat{s})$ be constructed from it by (4.3). Then*

$$(\hat{x}, \hat{y}, \hat{s}) \in \hat{\mathcal{N}}_2(\hat{\theta})$$

where $\hat{\theta} = \frac{n_2}{n_1}\theta + \sqrt{n_2}(1 - \rho^{(1,2)})$.

Proof. The slightly technical proof is given in the Appendix.

We also give a version of Lemma 4.1 using the $\mathcal{N}_{-\infty}$ neighbourhood (and ρ_-).

LEMMA 4.3. *Let $(x, y, s) \in \mathcal{N}_{-\infty}^{(1)}(\gamma)$ and let $(\hat{x}, \hat{y}, \hat{s})$ be constructed from it by (4.3). Then*

$$(\hat{x}, \hat{y}, \hat{s}) \in \hat{\mathcal{N}}_{-\infty}(\rho_-^{(1,2)}\gamma).$$

Proof. Since $x_i s_i \geq \gamma\mu$ we have

$$\hat{x}_i^t \hat{s}_i^t = x_i^{r(t)} s_i^{r(t)} \frac{p_t}{p_{r(t)}} \geq \frac{p_t}{p_{r(t)}}\gamma\mu = \frac{p_t}{p_{r(t)}}\gamma\frac{n_2}{n_1}\hat{\mu} \geq \rho_-^{(1,2)}\gamma\hat{\mu}.$$

□

4.2. Step 2: Bounds on δ_{bc} based on tree distance. The aim of this section is to relate the distance measure δ_{bc} to the distance between the old and new scenario trees. A similar result is given in [3, Lemma 7], but the bound here is tighter.

LEMMA 4.4. *Let $(\hat{x}, \hat{y}, \hat{s}) \in \hat{\mathcal{F}}^0$ be given. For the primal–dual infeasibilities incurred by changing the problem from $\hat{\mathcal{T}}$ to $\mathcal{T}^{(2)}$ we have*

$$\delta_{bc} \leq 2\sqrt{|\mathcal{T}^{(2)}|}\|W^T(WW^T)^{-1}\|_2 d(\mathcal{T}^{(1)}, \mathcal{T}^{(2)}).$$

Proof. According to Remark 1, δ_{bc} is equal to $\|\Delta x^*\|_2 + \|\Delta s^*\|_2$, where $\Delta x^*, \Delta s^*$ is the minimal change to $\hat{x}, \hat{s}$ needed to obtain a primal–dual feasible point for problem $\mathcal{T}^{(2)}$. Therefore for any change $\Delta x, \Delta s$ that restores primal–dual feasibility, $\|\Delta x\|_2 + \|\Delta s\|_2$ is an upper bound for δ_{bc}.

Since $(\hat{x}, \hat{y}, \hat{s}) \in \hat{\mathcal{F}}$ we have

$$p_t \hat{q}_t - W^T \hat{y}_t - \hat{s}_{2,t} = 0 \tag{4.4a}$$
$$c_1 - A^T \hat{y}_1 - T^T \hat{y}_{2,t} - \hat{s}_1 = 0 \tag{4.4b}$$
$$A\hat{x}_1 = b \tag{4.4c}$$
$$T\hat{x}_1 + W\hat{x}_{2,t} = \hat{h}_t \tag{4.4d}$$

We now consider the effect of changing the problem from $\hat{\mathcal{T}}$ to $\mathcal{T}^{(2)}$, that is replacing $\hat{q}_t$ and $\hat{h}_t$ by $q_t^{(2)}$ and $h_t^{(2)}$, respectively. Note that we have

$$\|\Delta q_t\|_2 = \|\hat{q}_t - q_t^{(2)}\|_2 \leq d(\mathcal{T}^{(1)}, \mathcal{T}^{(2)}), \qquad \|\Delta h_t\|_2 = \|\hat{h}_t - h_t^{(2)}\|_2 \leq d(\mathcal{T}^{(1)}, \mathcal{T}^{(2)}).$$

Since the change of problem data only affects q_t, h_t, equations (4.4b/c) are still satisfied. For (4.4a/d) we construct a change $\Delta x, \Delta s$ to restore primal–dual feasibility.

In order to satisfy (4.4a) we can simply take $\Delta s_{2,t} = p_t(q_t^{(2)} - \hat{q}_t)$ and get

$$p_t q_t^{(2)} - W^T \hat{y}_1 - (\hat{s}_{2,t} + \Delta s_{2,t}) = p_t \hat{q}_t - W^T \hat{y}_1 - \hat{s}_{2,t} = 0,$$

and therefore

$$\|\Delta s_{2,t}\| \leq p_t \|\Delta q_t\| \leq d(\mathcal{T}^{(1)}, \mathcal{T}^{(2)}).$$

For (4.4d) we look at finding a change $\Delta x_{2,t}$ to satisfy

$$\min \tfrac{1}{2}\Delta x_{2,t}^T \Delta x_{2,t} \text{ s.t. } T\hat{x}_1 + W(\hat{x}_{2,t} + \Delta x_{2,t}) = h_t^{(2)}.$$

This problem is feasible for every t due to the assumed full recourse. Deriving optimality conditions and using the fact that W has full rank, one can see easily that this is satisfied by

$$\Delta x_{2,t} = W^T(WW^T)^{-1}(h_t^{(2)} - \hat{h}_t),$$

hence

$$\|\Delta x_{2,t}\|_2 \leq \|W^T(WW^T)^{-1}\|_2 d(\mathcal{T}^{(1)}, \mathcal{T}^{(2)}).$$

Together we obtain

$$\begin{aligned}\delta_{bc} &\leq \|\Delta x\|_2 + \|\Delta s\|_2 \leq 2\sqrt{|\mathcal{T}^{(2)}|}\max\{\|\Delta x_{2,t}\|_2, \|\Delta s_{2,t}\|_2\} \\ &\leq 2\sqrt{|\mathcal{T}^{(2)}|}\|W^T(WW^T)^{-1}\|_2 d(\mathcal{T}^{(1)}, \mathcal{T}^{(2)}),\end{aligned}$$

where we have tacitly assumed that $\|W^T(WW^T)^{-1}\|_2 \geq 1$. □

4.3. Results for the two-tree scheme. We are now in a position to give conditions under which the whole warm-start process on two trees is successful. The algorithm we analyse consists of the steps detailed in Algorithm 1. In what follows we

Algorithm 1 Two-tree stochastic warm-start

Require: The full tree $\mathcal{T}_2$.

1: Starting from $\mathcal{T}_2$ choose a reduced tree $\mathcal{T}_1$ (for example by clustering and scenario selection).
2: Solve the problem on the reduced tree $\mathcal{T}_1$ to a duality gap of $\mu = \mu^{(1)}$, to obtain a point

$$(x^{(1)}, y^{(1)}, s^{(1)}) \in \mathcal{N}^{(1)}$$

for a specific choice of neighbourhood $\mathcal{N}^{(1)}$.
3: Expand this point by (4.3) to a solution $(\hat{x}, \hat{y}, \hat{s})$ of the expanded problem.
4: Do a modification step in the full problem.

give results for the $\mathcal{N}_2$ and $\mathcal{N}_{-\infty}$ neighbourhoods for the modification steps considered in Section 3.1. Unlike the related Theorem 8 from [3], these results give explicit dependencies of the conditions on μ, and problem specific data such as dimension n, while at the same time consider the amount of the necessary widening of the neighbourhood.

THEOREM 4.5. *Let $(x^{(1)}, y^{(1)}, s^{(1)}) \in \mathcal{N}^{(1)}_{-\infty}(\gamma_0)$ with $(x^{(1)})^T s^{(1)} = n_1\mu$. Then the warm-start according to Algorithm 1 is successful, that is*

$$(\hat{x} + \Delta x, \hat{y} + \Delta y, \hat{s} + \Delta s) \in \mathcal{N}^{(2)}_{-\infty}(\gamma),$$

where $(\Delta x, \Delta y, \Delta s)$ is a pure centring step in the full problem, if

$$d(\mathcal{T}_1, \mathcal{T}_2)\frac{n_2^{3/2}\sqrt{|\mathcal{T}_2|}}{n_1^{3/2}\rho_-} \leq \frac{\sqrt{\rho_-\gamma_0 - \gamma}}{2B_\infty^2\|W^T(WW^T)^{-1}\|_2}\gamma_0\mu^{3/2}.$$

Proof. According to Lemma 4.3, the expansion procedure of Step 3 leads to a point

$$(\hat{x}, \hat{y}, \hat{s}) \in \hat{\mathcal{N}}_{-\infty}(\rho_-\gamma_0), \text{ with } \hat{\mu} := \hat{x}^T\hat{s}/n_2 = \mu n_1/n_2.$$

According to Corollary 3.6 the warm-start is thus successful if

$$\delta_{bc} \leq \frac{\sqrt{\rho_-\gamma_0 - \gamma}}{2B_\infty^2}\rho_-\gamma_0(\frac{n_1}{n_2}\mu)^{3/2}.$$

Combining this with the bound on δ_{bc} from Lemma 4.4 we get the condition

$$\sqrt{|\mathcal{T}^{(2)}|}\|W^T(WW^T)^{-1}\|_2 d(\mathcal{T}^{(1)}, \mathcal{T}^{(2)}) \leq \frac{\sqrt{\rho_-\gamma_0 - \gamma}}{2B_\infty^2}\rho_-\gamma_0(\frac{n_1}{n_2}\mu)^{3/2},$$

which can be rearranged to obtain the condition of the Theorem. □

A result for the $\mathcal{N}_2$-neighbourhood is a bit harder to obtain, since in the worsening-of-centrality result (Lemma 4.2) the new centrality measure $\bar{\theta} = \frac{n_2}{n_1}\theta + \sqrt{n_2}(\rho - 1)$ depends on more parameters. We can however control the $\frac{n_2}{n_1}\theta$-term by reducing the

initial centrality measure through some additional centering steps. This leads to the following result.

THEOREM 4.6. *Let* $(x^{(1)}, y^{(1)}, s^{(1)}) \in \mathcal{N}_2^{(1)}(0.5)$ *with* $(x^{(1)})^T s^{(1)} = n_1\mu$. *Assume that*

$$n_2 \leq 10n_1, \quad 1 - \rho^{(1,2)} \leq \frac{1}{4\sqrt{n_2}}.$$

By performing up to two additional centering iterations in the reduced problem we can obtain a point

$$(\tilde{x}^{(1)}, \tilde{y}^{(1)}, \tilde{s}^{(1)}) \in \mathcal{N}_2(0.15\frac{n_1}{n_2}).$$

From this point the warm-start according to Algorithm 1 is successful, that is

$$(\hat{x} + \Delta x, \hat{y} + \Delta y, \hat{s} + \Delta s) \in \mathcal{N}_2^{(2)}(0.5),$$

where $(\Delta x, \Delta y, \Delta s)$ *is a pure centring step in the full problem, if*

$$d(\mathcal{T}_1, \mathcal{T}_2)\sqrt{|\mathcal{T}_2|}\frac{n_2^{5/4}}{n_1} \leq \frac{1}{104B_\infty\|W^T(WW^T)^{-1}\|_2}\mu.$$

REMARK 2. *These conditions are obviously very tight:*

$$1 - \rho^{(1,2)} \leq \frac{1}{4\sqrt{n_2}}$$

means that for moderate tree sizes ($n_2 \approx 10^5$) *we need* $\rho > 0.999$, *i.e. a more or less perfectly balanced tree.* In order to prove this Theorem we need the following generalisation of the centering results [9, Lemma 5.8]:

LEMMA 4.7. *Let* $(x, y, s) \in \mathcal{N}_2(\theta)$ *with* $0 < \theta < 0.5$ *and let* $(\Delta x, \Delta y, \Delta s)$ *be obtained by a pure centring step. Then the full step is feasible and*

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

where $\bar{\theta} = \frac{\theta^2}{2^{3/2}(1-\theta)}$.

Proof. By combining Lemmas 5.5/5.4 in [9] we get (setting $\alpha = \sigma = 1$):

$$\|(X + \Delta X)(S + \Delta S)e - \mu e\|_2 \leq \|\Delta X \Delta S e\|_2 \leq \frac{\theta^2}{2^{3/2}(1-\theta)}\mu.$$

The feasibility of the full step follows from [9, Lemma 5.8]. □

COROLLARY 4.8. *By performing a series of pure centring steps we have the progression*

$$\mathcal{N}_2(0.5) \to \mathcal{N}_2(0.18) \to \mathcal{N}_2(0.015) \to \mathcal{N}_2(0.001).$$

Proof (of Theorem 4.6). According to Corollary 4.8, a maximum of two centring steps will take us from $\mathcal{N}_2(0.5)$ to $\mathcal{N}_2(0.015) \subset \mathcal{N}_2(0.15\frac{n_1}{n_2})$ for $n_2 \leq 10n_1$. According to Lemma 4.2, the expansion procedure of Step 1 leads to a point

$$(\hat{x}, \hat{y}, \hat{s}) \in \hat{\mathcal{N}}_2(\hat{\theta}), \text{ with } \hat{\mu} := \hat{x}^T\hat{s}/n_2 = \mu n_1/n_2,$$

where

$$\hat{\theta} = \frac{n_2}{n_1}(0.15\frac{n_1}{n_2}) + \sqrt{n_2}(\rho - 1) \leq 0.15 + 0.25 = 0.4.$$

According to Corollary 3.4, the warm-start is thus successful if

$$\delta_{bc} \leq \frac{(1-\hat{\theta})^{3/2}(\theta - \hat{\theta})^{1/2}}{8\sqrt[4]{n_2}B_\infty}\frac{n_1}{n_2}\mu.$$

Combining this with the bound on δ_{bc} from Lemma 4.4 we get the condition

$$\sqrt{|\mathcal{T}^{(2)}|}\|W^T(WW^T)^{-1}\|_2 d(\mathcal{T}^{(1)}, \mathcal{T}^{(2)}) \leq \frac{(1-\hat{\theta})^{3/2}(\theta - \hat{\theta})^{1/2}}{8\sqrt[4]{n_2}B_\infty}\frac{n_1}{n_2}\mu.$$

Rearranging, setting $\theta = 0.5$ and using

$$\frac{(\theta - \hat{\theta})61/2}{8}(1-\hat{\theta})^{3/2} \geq \frac{1}{104},$$

we obtain the condition of the Theorem. □

5. Warm-start on a series of trees. The results of Section 4.3 give conditions under which the warm-start from a neighbourhood of the central path in the reduced tree leads to a neighbourhood of the central path in the full tree. In this section we apply those results to a series of progressively larger trees $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \cdots, \mathcal{T}^{(K)} = \mathcal{T}$. Algorithm 2 details the steps for the multiple tree stochastic warm-start scheme.

Algorithm 2 Multiple-tree stochastic programming warm-start

Require: A sequence of trees $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \cdots, \mathcal{T}^{(K)}$ with resulting problem sizes $n^{(k)}$ and closeness measures $d^{(k)} = d(\mathcal{T}^{(k)}, \mathcal{T}^{(k+1)})$. Further a sequence of μ-values $\mu^{(k)}$.

1: **for** $k = 1$ to K **do**
2: Solve problem $P(\mathcal{T}^{(k)})$ to accuracy $\mu^{(k)}$, that is to obtain a point $(x^{(k)}, y^{(k)}, s^{(k)}) \in \mathcal{N}^{(k)}$ for some neighbourhood $\mathcal{N}^{(k)}$ with $(x^{(k)})^T s^{(k)}/n^{(k)} = \mu^{(k)}$.
3: Perform additional centring steps to obtain a point in a tighter neighbourhood.
4: Expand this point to a solution $(\hat{x}^{(k)}, \hat{y}^{(k)}, \hat{s}^{(k)})$ of the expanded problem, by following the scheme in (4.3).
5: Use $(\hat{x}^{(k)}, \hat{y}^{(k)}, \hat{s}^{(k)})$ to warm-start the problem on the next tree in the sequence $\mathcal{T}^{(k+1)}$.
6: **end for**

In order to analyse the Algorithm we can use the results of Section 4. Their conditions, however, require the existence of a bound B_∞ on size of the components of the primal-dual iterate used for warmstarting. Such a bound can be derived by making use of the Renegar condition number of a problem and deriving a bound for it that is valid for all $P(\mathcal{T}^{(k)})$. Following the development in [10], we denote by $\mathcal{F}$ the set of primal–dual feasible LP data instances

$$\mathcal{F} := \{(A, b, c) : \exists (x, y, s) \text{ with } (x, s) > 0 \text{ such that } Ax = b, A^T y + s = c\},$$

its complement by $\mathcal{F}^C$, and their shared boundary as $\mathfrak{B} = \mathrm{cl}(\mathcal{F}) \cap \mathrm{cl}(\mathcal{F}^C)$. For a problem given by the data $\mathfrak{d} = (A, b, c)$, let

$$\|\mathfrak{d}\| := \max\{\|A\|_2, \|b\|_2, \|c\|_2\}$$

and the difference to ill-posedness $\rho(\mathfrak{d})$ be given by

$$\rho(d) = \inf\{\|\Delta d\| : d + \|\Delta d\| \in \mathfrak{B}\}. \tag{5.1}$$

The Renegar condition number $\mathcal{C}(\mathfrak{d})$ of the data instance $\mathfrak{d}$ is then defined as

$$\mathcal{C}(\mathfrak{d}) = \|\mathfrak{d}\|/\rho(\mathfrak{d}) \quad (\geq 1). \tag{5.2}$$

According to [7, Corollary 3.1] every primal–dual feasible point $(x^{(k)}, y^{(k)}, s^{(k)})$ with $\mu^{(k)} = (x^{(k)})^T s^{(k)}/n^{(k)}$ for a problem given by the data $\mathfrak{d}$ satisfies the bounds

$$\|x^{(k)}\|_\infty \leq \|x^{(k)}\|_1 \leq C(\mathfrak{d}^{(k)})^2 + n^{(k)}\mu^{(k)}/\rho(\mathfrak{d}^{(k)}) \tag{5.3a}$$

$$\|s^{(k)}\|_\infty \leq 2\|\mathfrak{d}^{(k)}\|_1[C(\mathfrak{d}^{(k)})^2 + n^{(k)}\mu^{(k)}/\rho(\mathfrak{d}^{(k)})]. \tag{5.3b}$$

Thus if we can derive globally valid bounds for the quantities $C(\mathfrak{d}^{(k)}), \rho(\mathfrak{d}^{(k)})$ and $\|\mathfrak{d}^{(k)}\|_1$ we can bound B_∞ using (5.3). The situation is summarised in the following Lemmas.

LEMMA 5.1. *Let $\mathcal{T}$ be the full tree and let the reduced tree $\overline{\mathcal{T}}$ be obtained from it by scenario aggregation. Then we have that*

$$\rho(\mathfrak{d}(\mathcal{T})) \leq \rho(\mathfrak{d}(\overline{\mathcal{T}})),$$

where $\mathfrak{d}(\mathcal{T}) = \mathfrak{d}(P(\mathcal{T}))$ is the problem data $\mathfrak{d} = (A, b, c)$ describing the problem $P(\mathcal{T})$.
Proof. The proof can be found in the Appendix.

LEMMA 5.2. *Given a sequence of trees $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \cdots, \mathcal{T}^{(K)}$ where $\mathcal{T}^{(k-1)}$ is obtained from $\mathcal{T}^{(k)}$ by scenario aggregation, then*

$$C(\mathfrak{d}^{(k)}) \leq \overline{C(\mathfrak{d})} = C(\mathfrak{d}^{(K)}), \quad \forall k.$$

Further, given a sequence of μ-values: $\mu^{(1)} \geq \mu^{(2)} \geq \ldots \geq \mu^{(K)}$, then with

$$\overline{B}_\infty = 2\|\mathfrak{d}^{(K)}\|_1(\overline{C(\mathfrak{d})}^2 + n^{(K)}\mu^{(1)}/\bar{\rho})$$

we have that all $(x^{(k)}, s^{(k)})$ that are used in the warm-start sequence satisfy

$$\|x^{(k)}\|_\infty, \|s^{(k)}\|_\infty \leq \overline{B}_\infty.$$

Proof. The existence of an upper bound $\overline{C(\mathfrak{d}^{(k)})} \leq \overline{C(\mathfrak{d})}$ follows from $C(\mathfrak{d}) = \|\mathfrak{d}\|/\rho(\mathfrak{d})$, $\|\mathfrak{d}^{(k)}\| \leq \|\mathfrak{d}^{(k+1)}\|$ and the existence of a lower bound $\bar{\rho} \leq \rho(\mathfrak{d}^{(k)})$ according to Lemma 5.1. Since $\|\mathfrak{d}^{(k)}\| \leq \|\mathfrak{d}^{(K)}\|, \mu^{(k)} \leq \mu^{(1)}$ and $n^{(k)} \leq n^{(K)}$ the assertion of the Lemma follows from inequalities (5.3a/b). □

We are now in a position to state the main results of this paper. We give conditions under which all the warm-start steps in the algorithm are successful both for the $\mathcal{N}_2$ and the $\mathcal{N}_{-\infty}$ neighbourhoods.

THEOREM 5.3. *Given is a sequence of trees $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \cdots, \mathcal{T}^{(K)}$ with resulting problem sizes $n^{(k)}$ and closeness measures $d^{(k)} = d(\mathcal{T}^{(k)}, \mathcal{T}^{(k+1)})$. Further we are*

given a sequence of μ-values $\mu^{(k)}$. Assume that we have a bound $\overline{C(\mathfrak{d})} = C(P^{(k)})$. If we have

$$n^{(k+1)} \leq 10n^{(k)}, \qquad \rho^{(k,k+1)} \leq 1 + \frac{1}{4\sqrt{n^{(k+1)}}}$$
$$\mu^{(k)} \geq 104\sqrt{|\mathcal{T}^{(k+1)}|}\bar{B}_\infty \|W^T(WW^T)^{-1}\|_2 d^{(k)} \frac{(n^{(k+1)})^{5/4}}{n^{(k)}}, \tag{5.4}$$

with $\bar{B}_\infty$ from Lemma 5.2, then Algorithm 2 can be performed successfully, that is all warm-starts are successful and lead to points

$$(x^{(k+1)}, y^{(k+1)}, s^{(k+1)}) \in \mathcal{N}_2^{(k+1)}(0.5).$$

Proof. All the warm-start points are primal–dual feasible for their respective originating problem. Therefore the bounds from Lemma 5.2 hold. The assertion of the Theorem then follows from Theorem 4.6.□

We will now give the equivalent result for the $\mathcal{N}_{-\infty}$ neighbourhood based on Theorem 4.5. In the $\mathcal{N}_{-\infty}$ neighbourhood there are no cheap re-centring steps. We therefore suggest a scheme in which the centrality measure γ worsens at every step. In particular we will choose the sequence $\gamma^{(0)} > 0.1, \gamma^{(k+1)} = (\gamma^{(k)} + 0.1)/2$.

THEOREM 5.4. *Given is a sequence of trees $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \cdots, \mathcal{T}^{(K)}$ with resulting problem sizes $n^{(k)}$ and closeness measures $d^{(k)} = d(\mathcal{T}^{(k)}, \mathcal{T}^{(k+1)})$. Further we are given a sequence of μ-values $\mu^{(k)}$. Assume that we have a bound $\overline{C(\mathfrak{d})} = C(P^{(k)})$. If we choose*

$$\gamma^{(k+1)} = \frac{1}{2}(\gamma^{(k)} + \frac{1}{10})$$

and satisfy the conditions

$$1 - \rho_-^{(k,k+1)} \leq \frac{1}{4}(\gamma^{(k)} - \frac{1}{10})$$
$$d^{(k)}\sqrt{|\mathcal{T}^{(k+1)}|} \leq \frac{3\sqrt{\gamma^{(k)} - \frac{1}{10}}}{2 \cdot 80\bar{B}_\infty^2 \|W^T(WW^T)^{-1}\|_2} \Big(\frac{n^{(k)}}{n^{(k+1)}}\mu\Big)^{3/2} \tag{5.5}$$

with $\bar{B}_\infty$ from Lemma 5.2, then Algorithm 2 can be performed successfully, that is all warm-starts are successful and lead to points

$$(x^{(k+1)}, y^{(k+1)}, s^{(k+1)}) \in \mathcal{N}_{-\infty}^{(k+1)}(\gamma^{(k+1)}).$$

Proof. All the warm-start points are primal–dual feasible for their respective originating problem, so the bounds from Lemma 5.2 hold. Under the conditions of the Theorem we have (since $\gamma < 1$)

$$1 - \rho \leq \frac{1}{4}(\gamma - \frac{1}{10}) \leq \frac{1}{4\gamma}(\gamma - \frac{1}{10}) = \frac{1}{4} - \frac{1}{40\gamma},$$

and therefore

$$\rho^{(k,k+1)} \geq \frac{3}{4} + \frac{1}{40\gamma^{(k)}}.$$

Moreover

$$\gamma^{(k)}\rho_- - \gamma^{(k+1)} \geq \frac{3}{4}\gamma^{(k)} + \frac{1}{40} - \frac{1}{2}\gamma^{(k)} - \frac{1}{20} = \frac{1}{4}(\gamma^{(k)} - \frac{1}{10}), \tag{5.6}$$

and therefore with $\rho_- \geq 3/4, \gamma^{(k)} \geq 1/10$:

$$\rho_-\sqrt{\rho_-\gamma^{(k)} - \gamma^{(k+1)}}\gamma^{(k)} \geq \frac{3}{4}\frac{1}{2}\sqrt{\gamma^{(k)} - \frac{1}{10}}\frac{1}{10} = \frac{3}{80}\sqrt{\gamma^{(k)} - \frac{1}{10}}. \tag{5.7}$$

Combining (5.7) with (5.5) we see that the condition of Theorem 4.5 is satisfied. □

5.1. Complexity analysis. We now take a closer look at Algorithm 2. In step 2 the advanced $\mu^{(k)}$-centre is found by applying an interior point method. Assuming (as usual) that we know a primal–dual feasible μ^0 centre, the initial $\mu^{(1)}$-centre can be found in

$$\mathcal{O}\left(n^\tau \log \frac{\mu^0}{\mu^{(1)}}\right) \tag{5.8}$$

iterations (see for example [10]). The parameter τ depends on the IPM variant used, and is $\tau = 1$ for the long-step method and $\tau = 1/2$ for the short-step method. We assume that the computational cost of an interior point iteration (which is dominated by the solution of system (2.8)) is $\mathcal{O}(n^\alpha)$, where $\alpha > 1$ (dense linear algebra would give $\alpha = 3$, the best structure exploiting solvers are able to reach $\alpha \approx 1$ in specific circumstances).

Since typically $\alpha > 1$, the computational cost will be dominated by the number of iterations that have to be performed on the full problem $P(\mathcal{T})$. It is therefore crucial to analyse what the latest point (smallest μ) is at which the switch to the full problem can be performed successfully according to the analysis presented in the previous sections.

If we use only one reduced tree for warm-starting (as in [3]), then from condition (5.4) of Theorem 5.3 and the fact that typically $n^{(k)} \approx C|\mathcal{T}^{(k)}|$ we get as a condition for the smallest acceptable μ-value (in the reduced problem) the bound

$$\underline{\mu^{(1)}} \geq Cd(\mathcal{T}_1, \mathcal{T}^{(k)})n_K^{7/4}/n_1.$$

Similar conditions can be obtained using condition (5.5) to obtain bounds of the form

$$\underline{\mu^{(1)}} \geq Cd(\mathcal{T}_1, \mathcal{T}^{(k)})^a n_K^b/n_1,$$

where the values of a, b and the constant C depend on which strand of the analysis we take according to

cond	(5.4)	(5.5)
a	1	2/3
b	7/4	4/3

According to the analysis of the expansion and warm-start steps this yields a warm-start point in the full problem with a corresponding μ-value

$$\underline{\mu} = \underline{\mu^{(1)}}\frac{n_1}{n_K} \geq Cd(\mathcal{T}_1, \mathcal{T}^{(k)})^a n_K^{b-1}. \tag{5.9}$$

One the other hand if we use a sequence of trees $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}, \mathcal{T}^{(K)}$, the final warm-start into the full problem $P(\mathcal{T}^{(K)})$ will be performed from $\mathcal{T}^{(2)}$ and by a similar argument we obtain the bound

$$\underline{\mu} = \underline{\mu^{(2)}}\frac{n_2}{n_K} \geq Cd(\mathcal{T}_2, \mathcal{T}^{(k)})^a n_K^{b-1}. \tag{5.10}$$

We expect that $\mathcal{T}^{(2)}$ is chosen such that $d(\mathcal{T}^{(2)}, \mathcal{T}^{(K)}) < d(\mathcal{T}^{(1)}, \mathcal{T}^{(K)})$. Comparing (5.9) and (5.10) and using (5.8) we obtain that the number of full problem iterations saved by performing the multi-step warm-start is

$$\mathcal{O}\left(a \log \frac{d(\mathcal{T}^{(2)}, \mathcal{T}^{(K)})}{d(\mathcal{T}^{(1)}, \mathcal{T}^{(K)})}\right).$$

A similar analysis can now be applied to determine whether it is worthwhile to introduce an additional tree $\mathcal{T}^{(3)}$ between $\mathcal{T}^{(2)}$ and $\mathcal{T}^{(K)}$, or indeed whether problem $P(\mathcal{T}^{(2)})$ should be warm-started directly from $\mathcal{T}^{(1)}$ or from an intermediate tree between $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$.

We should note that typically there is a large gap between the theoretically predicted and practically observed behaviour of interior point methods. We therefore see the bounds in Theorems 5.3 and 5.4 only as qualitative guidance, but will not check them in practice. We think this reasoning is supported by the numerical results which we present in the final section.

6. Numerical results. The multi-step warm-start algorithm is implemented within OOPS [6], an infeasible primal–dual long-step interior point code. In fact, OOPS does not enforce that iterates have to be in a specific neighbourhood. As such the code is quite different from the exemplary method used in our analysis. Our aim is however to demonstrate that the warm-starting scheme offers potential for time savings even with production codes (and not just on a textbook algorithm).

We have tested the multi-step warm-start algorithm (Algorithm 2) on two sets of test problems. The first set consists of variations of the standard diet problem in which a set of products have to be mixed from a supply of raw materials in order to satisfy certain nutrient specifications of the products at minimum cost. The problem is made stochastic by assuming that demand of products and availability of raw materials are unknown. The first-stage decision is to decide how much to buy of each raw material, and the recourse problem decides how to mix each of the products. We allow to buy in missing raw materials at premium cost to ensure full recourse. The second set comes from capacity assignment problems with uncertain demand. Scenarios for both sets of problems are generated by sampling from a (multivariate) uniform distribution. Details of our test problems and their dimensions are given in Table 6.1.

		core size		problem size	
Problem	scenarios	constraints	variables	constraints	variables
ex1	20000	50	70	960.001	1.780.006
	40000	50	70	1.920.001	3.560.006
ex3	10000	72	112	700.001	1.380.009
	20000	72	112	1.400.001	2.760.009
s97	10000	147	157	1.450.001	2.390.010
s98	10000	149	141	1.470.001	2.060.006
j99	10000	158	148	1.560.001	2.170.007
Minoux	10000	118	239	1.160.001	3.050.051
Jll_gva	4000	350	781	1.392.001	4.180.085
T1B3	10000	50	109	1.070.001	1.350.024
r4c	10000	177	272	2.700.001	2.970.151

TABLE 6.1
Dimensions of the test problems.

All computations have been performed on a Dual Core Intel 3.0GHz processor with 4GB RAM. Starting from the full problem we generate reduced problems by clustering scenarios (using k-means++ clustering [1]). In our first set of results (presented in Table 6.2) we compare the efficiency of a single step of the stochastic warm-start (Algorithm 1) for varying sizes of the reduced tree. The numbers given are the IPM iterations for the full tree (after warm-start), the total solution time (for the reduced and full problem solve) and the combined residual $\|\xi_b\|_\infty + \|\xi_c\|_\infty$ in the first iteration of the full problem.

As expected, the residuals in the full problem decrease as the reduced tree approximation improves, confirming the assertions of Lemma 4.4. In all cases, the warm-start can significantly reduce the number of iterations. Since an iteration in the reduced problem is much cheaper than an iteration in the full problem, the savings in iterations are translated into savings of solution time. As expected the number of iterations needed in the warm-started problem is generally smaller if the reduced tree is larger, i.e. it approximates the full tree better. However, in terms of solution time there is a slight trade-off, since using a large reduced tree means that a non-negligible time is spent in finding the warm-start point.

To complete the picture, Table 6.3 gives the smallest target μ-value in the reduced tree for which the warm-start is successful found by experimentation. Since we deem the definition of a "successful warm-start" used in the theoretical part of this paper (namely the modification step itself regains feasibility in the full problem) too stringent, we have used the criterion that the infeasible IPM used for the full problem is able to reduce the combined primal–dual infeasibility below 0.1 or 1/100 of its original value, whichever is smaller, within the first 3 iterations. As can be seen, the smallest successful μ-value decreases as the quality of the reduced tree approximation improves, again confirming the results of Theorems 4.5 and 4.6.

In our second set of results we test the performance of the multi-step scheme. In all instances we have used three trees: for each problem the full tree and the smallest and largest reduced tree from Table 6.2. In Table 6.4 we compare the solution time for the cold started problem, the problem warm-started from the largest reduced tree, problem warm-started from the smallest reduced tree and finally the three-tree scheme. We also report the percentual time savings of the 3-step scheme compared to the best of the 2-step warm-starts (negative values indicate that the 3-step scheme is not the fastest). In all cases the multistep warm-start can improve on the solution time of the of the 2-step scheme when warm-started from the largest tree. In most cases we can also improve on the solution time when warm-starting from the smallest reduced tree. Although the 3-step scheme is not in all cases the fastest of the warm-start schemes, it is always good. Keeping in mind that the optimal size of the reduced tree for a 2-step scheme can not be determined a-priori, the 3-step warm-start is a good alternative.

Our numerical results show that the multi-step warm-start scheme significantly speeds up the solution of large two-stage stochastic programming problems when compared to a cold start and has advantages compared to a simple reduced tree warm-start. Although the analysis has been performed for linear two-stage problems, we would expect that the scheme can be adapted to nonlinear problems and multi-stage stochastic programming. We leave this for future research.

REFERENCES

Problem	scenarios	cold	size of reduced tree				
			4000	2000	1000	100	10
ex1	20000	25	-	12	-	14	12
		580s		327s		311s	264s
				1.76		4.20	8.34
	40000	34	14	-	-	14	18
		1559s	766s			642s	843s
			1.55			3.87	7.40
ex3	10000	34	-	-	18	28	27
		563s			346s	469s	444s
					2.90	6.16	13.2
	20000	53	-	16	-	24	29
		1793s		626s		804s	961s
				2.56		5.75	11.7
s97	10000	31	-	-	15	18	15
		498s			389s	252s	229s
					2.46	4.26	7.88
s98	10000	>100	-	-	18	32	30
		2278s			826s	951s	878s
					7.77	11.8	23.2
j99	10000	76	-	-	11	11	11
		1796s			375s	282s	275s
					4.70	11.0	18.0
Minoux	10000	40	-	-	17	30	27
		2644s			1212s	1831	1790s
					0.081	0.441	3.52
Jll_gva	4000	63	-	-	30	25	44
		4981s			2523s	2284	2995s
					0.14	1.00	6.84
T1B3	10000	57	-	-	31	40	31
		995s			633s	679	557
					1.21	1.60	1.91
r4c	10000	41	-	-	14	21	24
		2098			835	1099	1204
					0.36	0.83	2.96

TABLE 6.2
Warm-started iterations, solution time and residuals for different sizes of the reduced tree.

[1] D. ARTHUR AND S. VASSILVITSKH, *k-means++: The advantages of careful seeding*, in Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms, 2007, pp. 1027–1035.

[2] H. Y. BENSON AND D. F. SHANNO, *An exact primal-dual penalty method approach to warmstarting interior-point methods for linear programming*, Comput Optim Appl, 38 (2007), pp. 371–399.

[3] M. COLOMBO, J. GONDZIO, AND A. GROTHEY, *A warm-start approach for large-scale stochastic linear programs*, Math Program, (2009). published online 30 May 2009.

[4] J. GONDZIO AND A. GROTHEY, *Reoptimization with the primal-dual interior point method*, SIAM J Opimiz, 13 (2003), pp. 842–864.

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

[6] ———, *Exploiting structure in parallel implementation of interior point methods for optimization*, Comput Manage Sci, 6 (2009), pp. 135–160.

[7] M. A. NUNEZ AND R. M. FREUND, *Condition measures and properties of the central trajectory*, Math Program, 83 (1998), pp. 1–28.

Problem	scenarios	4000	2000	1000	100	10
ex1	20000	-	1.5e-3	-	5e-2	1
	40000	5e-4	-	-	5e-2	1
ex3	10000	-	-	0.2	1	30
	20000	-	0.1	-	2	50
s97	10000	-	-	1e-4	2e-3	5e-2
s98	10000	-	-	0.2	1	50
j99	10000	-	-	1e-2	2	100
Minoux	10000	-	-	1.5e-4	2.e-3	1.5e-2
Jll_gva	4000	-	-	1.5e-4	1.5e-3	1.5e-2
T1B3	10000	-	-	1.e-4	1.e-3	1.e-2
r4c	10000	-	-	5.e-5	1.e-3	2.e-2

TABLE 6.3
Smallest μ for successful warm-start.

Problem	scenarios	cold	2-step (large)	2-step (small)	3-step	impr
ex1	20000	580	327	264	302	-14.4
	40000	1559	766	642	701	-9.2
ex3	10000	563	346	444	316	+8.7
	20000	1793	626	961	586	+6.4
s97	10000	498	389	229	307	-34.1
s98	10000	3189	826	878	481	+41.8
j99	10000	1796	375	275	295	-7.3
Minoux	10000	2644	1212	1790	1176	+3.0
Jll_gva	4000	4981	2523	2995	2251	+10.8
T1B3	10000	995	663	557	637	-14.4
r4c	10000	2098	835	1204	749	+10.3

TABLE 6.4
Comparison of solution time for cold start, 2-step warm-start and multi-step warm-start.

[8] A. SHAPIRO, D. DENTCHEVA, AND A. RUSZCZYNSKI, *Lecture Notes on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.

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

[10] E. A. YILDIRIM AND S. J. WRIGHT, *Warm-start strategies in interior-point methods for linear programming*, SIAM J Opimiz, 12 (2002), pp. 782–810.

Appendix A. *Proof of Lemma 4.2.* According to the proof of [3, Theorem 3], $(\hat{x}, \hat{y}, \hat{s})$ is primal–dual feasible for problem $P(\hat{\mathcal{T}})$. Also from the same proof we have

$$\hat{\mu} := \hat{x}^T \hat{s}/n_2 = n_1\mu_1/n_2.$$

We need to prove a bound on $\|\hat{X}\hat{S}e - \hat{\mu}e\|_2$. $(\hat{x}, \hat{s})$ as constructed by (4.3), are such that for a perfectly centred x, s, i.e. $x_i s_i = \mu_1$, we have

$$\hat{x}_i \hat{s}_i = \tilde{\mu}_i$$

where

$$\tilde{\mu}_i^{(t)} = \frac{p_t}{p_{r(t)}}\mu_1.$$

Here and in what follows we use the notation that $\tilde{\mu}_i^{(t)}$ is the i-th component of the part of $\tilde{\mu}$ corresponding to node $t \in \mathcal{T}_1$.

We have

$$\|\hat{X}\hat{S}e - \hat{\mu}e\|_2 \leq \|\hat{X}\hat{S}e - \tilde{\mu}\|_2 + \|\tilde{\mu} - \hat{\mu}e\|_2.$$

We will now derive bounds on these two terms separately:

$$\begin{aligned}
\|\hat{X}\hat{S}e - \tilde{\mu}\|_2^2 &= \sum_{s\in\mathcal{T}_1}\sum_{t\in\mathcal{I}(s)}\sum_i (\hat{x}_i^{(t)}\hat{s}_i^{(t)} - \hat{\mu}_i^{(t)})^2 \\
&= \sum_{s\in\mathcal{T}_1}\sum_{t\in\mathcal{I}(s)}\sum_i (\frac{p_t}{p_s}x_i^{(s)}s_i^{(s)} - \frac{p_t}{p_s}\mu_1)^2 \\
&\leq \sum_{s\in\mathcal{T}_1}\sum_{t\in\mathcal{I}(s)}\sum_i \frac{p_t}{p_s}(x_i^{(s)}s_i^{(s)} - \mu_1)^2 \\
&= \sum_{s\in\mathcal{T}_1}\frac{1}{p_s}\sum_i (x_i^{(s)}s_i^{(s)} - \mu_1)^2 \sum_{t\in\mathcal{I}(s)} p_t \\
&= \sum_{s\in\mathcal{T}_1}\sum_i (x_i^{(s)}s_i^{(s)} - \mu_1)^2 = \|XSe - \mu_1 e\|_2^2 \leq \theta^2\mu_1^2,
\end{aligned}$$

and

$$\begin{aligned}
\|\mu_1 e - \tilde{\mu}\|_2^2 &= \sum_{s\in\mathcal{T}_1}\sum_{t\in\mathcal{I}(s)}\sum_i (\tilde{\mu}_i^{(t)} - \hat{\mu})^2 \\
&= \sum_{s\in\mathcal{T}_1}\sum_{t\in\mathcal{I}(s)}\sum_i \mu_1^2(\frac{p_t}{p_s} - \frac{n_1}{n_2})^2 \\
&\leq \sum_{s\in\mathcal{T}_1}\sum_{t\in\mathcal{I}(s)}\sum_i \mu_1^2\frac{n_1^2}{n_2^2}(\rho - 1)^2 \\
&= \mu_1^2\frac{n_1^2}{n_2}(\rho - 1)^2,
\end{aligned}$$

so that by combining the two we have

$$\|\hat{X}\hat{S}e - \hat{\mu}e\|_2 \leq \theta\mu_1 + \frac{n_1}{\sqrt{n_2}}(1-\rho)\mu_1 = \theta\frac{n_2}{n_1}\hat{\mu} + \sqrt{n_2}(1\rho)\hat{\mu},$$

from which the assertion follows. □

Proof of Lemma 5.1. Let $\hat{\rho} = \rho(\mathfrak{d}(\mathcal{T}))$. Further let $\bar{\mathfrak{d}} = \mathfrak{d}(\overline{\mathcal{T}})$. What we need to show is that for any perturbation $\overline{\Delta\mathfrak{d}}$ that satisfies $\|\overline{\Delta\mathfrak{d}}\| \leq \hat{\rho}$ we have

$$\bar{\mathfrak{d}} + \overline{\Delta\mathfrak{d}} \in \mathcal{F}.$$

For what follows we use the shorthand $n = |\mathcal{T}|$, $\bar{n} = |\overline{\mathcal{T}}|$. With this problem $\mathfrak{d} = P(\mathcal{T})$ is

$$\begin{array}{lll}
\min & c_0^T x_0 + \sum_{i=1}^n p_i c_i^T x_i & \\
\text{s.t.} & W_0 x_0 & = h_0 \\
& p_i T x_0 + p_i W x_i & = p_i h_i, \quad i = 1, \ldots, n
\end{array}$$

Let us write problem $\bar{\mathfrak{d}} = P(\overline{\mathcal{T}})$ that is obtained by an aggregation of $\mathcal{T}$ in the following

form

$$\begin{array}{ll} \min & \bar{c}_0^T x_0 + \sum_{i=1}^{\bar{n}} \bar{p}_i \bar{c}_i^T x_i \\ \text{s.t.} & \begin{bmatrix} \bar{W}_0 & 0 & \cdots & 0 \\ \bar{p}_1 \bar{T} & \bar{p}_1 \bar{W} & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ \bar{p}_{\bar{n}} \bar{T} & 0 & 0 & \bar{p}_{\bar{n}} \bar{W} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{\bar{n}} \end{bmatrix} = \begin{bmatrix} \bar{h}_0 \\ \bar{p}_1 \bar{h}_1 \\ \vdots \\ p_{\bar{n}} \bar{h}_{\bar{n}} \end{bmatrix} \end{array}$$

where

$$\begin{array}{l} \bar{W}_0 = W_0, \quad \bar{W} = W, \quad \bar{T} = T, \quad \bar{h}_0 = h_0, \quad \bar{c}_0 = c_0, \\ \bar{p}_i = \displaystyle\sum_{j \in I(i)} p_j, \quad \bar{c}_i = (\sum_{j \in I(i)} p_j c_j)/\bar{p}_i, \quad \bar{h}_i = (\sum_{j \in I(i)} p_j h_j)/\bar{p}_i. \end{array}$$

Let us know consider the perturbed problem $\bar{\mathfrak{d}} + \overline{\Delta \mathfrak{d}}$ written as

$$\begin{array}{ll} \min & ([\bar{c}_0, \bar{p}_1 \bar{c}_1, \ldots, \bar{p}_{\bar{n}} \bar{c}_{\bar{n}}]^T + \overline{\Delta c})^T x \\ \text{s.t.} & \left(\begin{bmatrix} \bar{W}_0 & 0 & \cdots & 0 \\ \bar{p}_1 \bar{T} & \bar{p}_1 \bar{W} & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ \bar{p}_{\bar{n}} \bar{T} & 0 & 0 & \bar{p}_{\bar{n}} \bar{W} \end{bmatrix} + \overline{\Delta A} \right) \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{\bar{n}} \end{bmatrix} = \begin{bmatrix} \bar{h}_0 \\ \bar{p}_1 \bar{h}_1 \\ \vdots \\ p_{\bar{n}} \bar{h}_{\bar{n}} \end{bmatrix} + \overline{\Delta b} \end{array}$$

with

$$\overline{\Delta A} = \begin{bmatrix} \overline{\Delta W_0} & \overline{\Delta A_{0,1}} & \cdots & \overline{\Delta A_{0,\bar{n}}} \\ \bar{p}_1 \overline{\Delta T_1} & \bar{p}_1 \overline{\Delta W_1} & \overline{\Delta A_{1,i}} & \overline{\Delta A_{1,\bar{n}}} \\ \vdots & \overline{\Delta A_{i,1}} & \ddots & \vdots \\ \bar{p}_{\bar{n}} \overline{\Delta T_{\bar{n}}} & \overline{\Delta A_{\bar{n},1}} & \overline{\Delta A_{\bar{n},i}} & \bar{p}_{\bar{n}} \overline{\Delta W_{\bar{n}}} \end{bmatrix}, \quad \overline{\Delta b} = \begin{bmatrix} \overline{\Delta h_0} \\ \bar{p}_1 \overline{\Delta h_1} \\ \vdots \\ \bar{p}_{\bar{n}} \overline{\Delta h_{\bar{n}}} \end{bmatrix}, \quad \overline{\Delta c} = \begin{bmatrix} \overline{\Delta c_0} \\ \bar{p}_1 \overline{\Delta c_1} \\ \vdots \\ \bar{p}_{\bar{n}} \overline{\Delta c_{\bar{n}}} \end{bmatrix}$$

where we assume that $\|\Delta \mathfrak{d}\| \leq \hat{\rho}$. The perturbation $\overline{\Delta \mathfrak{d}}$ is expanded into a perturbation $\Delta \mathfrak{d} = (\Delta A, \Delta b, \Delta c)$ of the full problem

$$\Delta A = \begin{bmatrix} \Delta A_{0,0} & \Delta A_{0,1} & \cdots & \Delta A_{0,n} \\ \Delta A_{1,0} & \Delta A_{1,1} & \cdots & \Delta A_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta A_{n,0} & \Delta A_{n,1} & \cdots & \Delta A_{n,n} \end{bmatrix}, \quad \Delta b = \begin{bmatrix} \Delta b_0 \\ \Delta b_1 \\ \vdots \\ \Delta b_n \end{bmatrix}, \quad \Delta c = \begin{bmatrix} \Delta c_0 \\ \Delta c_1 \\ \vdots \\ \Delta c_n \end{bmatrix}$$

according to

$$\begin{array}{rcll} \Delta b_0 & = & \overline{\Delta h_0}, \quad \Delta b_i = p_i \overline{\Delta h_{r(i)}} & \\ \Delta c_0 & = & \overline{\Delta c_0}, \quad \Delta c_i = p_i \overline{\Delta c_{r(i)}} & \\ \Delta A_{0,0} & = & \overline{\Delta W_0}, \quad \Delta A_{i,i} = p_i \overline{\Delta W_{r(i)}} & \\ \Delta A_{0,i} & = & p_i \overline{\Delta A_{0,r(i)}}/\bar{p}_{r(i)}, & i \in \mathcal{T} \setminus \{0\} \\ \Delta A_{i,0} & = & p_i \overline{\Delta T_{r(i)}}, & i \in \mathcal{T} \setminus \{0\} \\ \Delta A_{i,j} & = & p_i p_j \overline{\Delta A_{r(i),r(j)}}/(\bar{p}_{r(i)} \bar{p}_{r(j)}), & i,j \in \mathcal{T} \setminus \{0\}, r(i) \neq r(j) \\ \Delta A_{i,j} & = & 0, & i,j \in \mathcal{T} \setminus \{0\}, r(i) = r(j), i \neq j \end{array}$$

Since $\|\overline{\Delta \mathfrak{d}}\| \leq \hat{\rho}$ we have

$$\|\overline{\Delta c}\|_2, \|\overline{\Delta b}\|_2, \|\overline{\Delta A}\|_2 \leq \hat{\rho}.$$

From $\bar{p}_j = \sum_{i\in I(j)} p_i$ we get $\sum_{i\in I(j)} p_i^2 \le \bar{p}_j^2$ and hence

$$\begin{aligned}\|\Delta c\|_2^2 &= \|\Delta c_0\|_2^2 + \sum_{i\in\mathcal{T}} \|\Delta c_i\|_2^2 = \|\overline{\Delta c}_0\|_2^2 + \sum_{j\in\overline{\mathcal{T}}} \sum_{i\in I(j)} p_i^2 \|\overline{\Delta c}_j\|_2^2 \\ &\le \|\overline{\Delta c}_0\|_2^2 + \sum_{j\in\overline{\mathcal{T}}} \bar{p}_j^2 \|\overline{\Delta c}_j\|_2^2 = \|\overline{\Delta c}\|_2^2 \le \hat{\rho}^2.\end{aligned}$$

Similarly, we obtain $\|\Delta b\|_2 \le \hat{\rho}$ and $\|\Delta A\|_2 \le \hat{\rho}$. Therefore we also have that $\|\Delta\mathfrak{d}\| \le \hat{\rho}$ and thus $\mathfrak{d} + \Delta\mathfrak{d} \in \mathcal{F}$, which means that there is a primal–dual feasible point (x, y, s) for $\mathfrak{d} + \Delta\mathfrak{d}$:

$$(W_0 + \Delta A_{0,0})x_0 + \sum_{i=1}^{n} \Delta A_{0,i} x_i = h_0 + \Delta b_0 \tag{A.1a}$$

$$(p_i T + \Delta A_{i,0})x_0 + (p_i W + \Delta A_{i,i} x_i) + \sum_{k\neq i} \Delta A_{i,k} x_k = p_i h_i + \Delta b_i \tag{A.1b}$$

$$(W_0 + \Delta A_{0,0})^T y_0 + \sum_{i=1}^{n} (p_i T + \Delta A_{i,0})^T y_i + s_0 = c_0 + \Delta c_0 \tag{A.1c}$$

$$(p_i W + \Delta A_{i,i})^T y_i + \sum_{k\neq i} \Delta A_{k,i}^T y_k + s_i = p_i c_i + \Delta c_i \tag{A.1d}$$

We aim to show that with the aggregations

$$\begin{array}{rclrclrcl} \bar{x}_0 &=& x_0 & \bar{y}_0 &=& y_0, & \bar{s}_0 &=& s_0, \\ \bar{x}_i &=& (\sum_{j\in I(i)} p_j x_j)/\bar{p}_i, & \bar{y}_i &=& (\sum_{j\in I(i)} p_j y_j)/\bar{p}_i, & \bar{s}_i &=& \sum_{j\in I(i)} s_j, \end{array}$$

the point $(\bar{x}, \bar{y}, \bar{s})$ is primal–dual feasible for $\bar{\mathfrak{d}} + \overline{\Delta\mathfrak{d}}$. We have

$$\begin{aligned}\sum_{i\in\mathcal{T}} \Delta A_{0,i} x_i &= \sum_{j\in\overline{\mathcal{T}}} \sum_{i\in I(j)} \Delta A_{0,i} x_i = \sum_{j\in\overline{\mathcal{T}}} \sum_{i\in I(j)} p_i \overline{\Delta A}_{0,j} / \bar{p}_j x_i \\ &= \sum_{j\in\overline{\mathcal{T}}} \overline{\Delta A}_{0,j} \Big(\sum_{i\in I(j)} p_i x_i \Big)/\bar{p}_j = \sum_{j\in\overline{\mathcal{T}}} \overline{\Delta A}_{0,j} \bar{x}_j,\end{aligned}$$

so that from (A.1a) we get

$$\begin{aligned}(\bar{W}_0 + \overline{\Delta W_0})\bar{x}_0 + \textstyle\sum_{j\in\overline{\mathcal{T}}} \overline{\Delta A}_{0,j} \bar{x}_j \quad &= (W_0 + \Delta A_{0,0})x_0 + \textstyle\sum_{i\in\mathcal{T}} \Delta A_{0,i} x_i \\ &= h_0 + \Delta b_0 = \bar{h}_0 + \overline{\Delta h}_0.\end{aligned} \tag{A.2}$$

We will continue by summing up (A.1b) for all $i \in I(j)$. From the $\sum_{k\neq i} \Delta A_{i,k} x_k$ term we get thus

$$\begin{aligned}\sum_{i\in I(j)} \sum_{k\neq i} \Delta A_{i,k} x_k &= \sum_{i\in I(j)} \sum_{k\neq i : r(k)\neq j} p_i p_k \overline{\Delta A}_{j,r(k)} / (\bar{p}_j \bar{p}_{r(k)}) x_k \\ &= \sum_{i\in I(j)} \sum_{l\in\overline{\mathcal{T}} : l\neq j} \sum_{k\in I(l)} p_i p_k \overline{\Delta A}_{j,l} / (\bar{p}_j \bar{p}_l) x_k \\ &= \sum_{l\in\overline{\mathcal{T}} : l\neq j} \overline{\Delta A}_{j,l} / \bar{p}_l \sum_{k\in I(l)} p_k x_k = \sum_{l\in\overline{\mathcal{T}} : l\neq j} \overline{\Delta A}_{j,l} \bar{x}_l\end{aligned}$$

So that summing up (A.1b) for all $i \in I(j)$ we get (using $\bar{p}_j = \sum_{i \in I(j)} p_i$):

$$\bar{p}_j(T + \overline{T}_j)x_0 + (W + \overline{\Delta W}_j)(\sum_{i \in I(j)} p_i x_i) + \sum_{l \in \overline{\mathcal{T}}: l \neq j} \overline{\Delta A}_{j,l}\bar{x}_l = \bar{p}_j \bar{h}_j + \bar{p}_j \overline{\Delta h}_j$$

which yields

$$\bar{p}_j(T + \overline{T}_j)\bar{x}_0 + \bar{p}_j(W + \overline{\Delta W}_j)\bar{x}_j + \sum_{l \in \overline{\mathcal{T}}: l \neq j} \overline{\Delta A}_{j,l}\bar{x}_l = \bar{p}_j(\bar{h}_j + \overline{\Delta h}_j). \tag{A.3}$$

Equations (A.2) and (A.3) together give the primal feasibility of $(\bar{x}, \bar{y}, \bar{s})$ in $\bar{\mathfrak{d}} + \overline{\Delta \mathfrak{d}}$. In the same manner as above we can write

$$\sum_{i \in \mathcal{T}} \Delta A_{i,0}^T y_i = \sum_{j \in \overline{\mathcal{T}}} \sum_{i \in I(j)} p_i \overline{\Delta T}_j^T y_i = \sum_{j \in \overline{\mathcal{T}}} \bar{p}_j \overline{\Delta T}_j^T \bar{y}_j,$$

so that from (A.1c) we get

$$(W_0 + \overline{\Delta W}_0)\bar{y}_0 + \sum_{j \in \overline{\mathcal{T}}} \bar{p}_j (T + \overline{\Delta T}_j)^T \bar{y}_j + s_0 = c_0 + \Delta c_0. \tag{A.4}$$

Finally, we sum up (A.1d) over $i \in I(j)$. From the $\sum_{k \neq i} \Delta A_{k,i}^T y_k$ term we get

$$\begin{aligned}
\sum_{i \in I(j)} \sum_{k \neq i} \Delta A_{k,i}^T y_k &= \sum_{i \in I(j)} \sum_{k \neq i: r(k) \neq j} p_k p_i \overline{\Delta A}_{r(k),j}^T / (\bar{p}_{r(k)} \bar{p}_j) y_k \\
&= \sum_{i \in I(j)} \sum_{l \in \overline{\mathcal{T}}: l \neq j} \sum_{k \in I(l)} p_k p_i \overline{\Delta A}_{l,j}^T / (\bar{p}_l \bar{p}_j) y_k \\
&= \sum_{l \in \overline{\mathcal{T}}: l \neq j} \overline{\Delta A}_{l,j}^T / \bar{p}_l \sum_{k \in I(l)} p_k y_k = \sum_{l \in \overline{\mathcal{T}}: l \neq j} \overline{\Delta A}_{l,j}^T \bar{y}_l
\end{aligned}$$

So that summing up (A.1d) for all $i \in I(j)$ we get (using $\bar{p}_j = \sum_{i \in I(j)} p_i$):

$$(W + \overline{\Delta W}_j)^T (\sum_{i \in I(j)} p_i y_i) + \sum_{l \in \overline{\mathcal{T}}: l \neq j} \overline{\Delta A}_{l,j}^T \bar{y}_l + \sum_{i \in I(j)} s_i = \bar{p}_j \bar{c}_j + \bar{p}_j \overline{\Delta c}_j$$

which yields

$$\bar{p}_j(W + \overline{\Delta W}_j)\bar{y}_j + \sum_{l \in \overline{\mathcal{T}}: l \neq j} \overline{\Delta A}_{l,j}^T \bar{y}_l + \bar{s}_j = \bar{p}_j(\bar{c}_j + \overline{\Delta c}_j). \tag{A.5}$$

Equations (A.4) and (A.5) together give the dual feasibility of $(\bar{x}, \bar{y}, \bar{s})$ in $\bar{\mathfrak{d}} + \overline{\Delta \mathfrak{d}}$. Together $\bar{\mathfrak{d}} + \overline{\Delta \mathfrak{d}}$ is primal–dual feasible, which gives $\bar{\mathfrak{d}} + \overline{\Delta \mathfrak{d}} \in \mathcal{F}$. Since $\overline{\Delta \mathfrak{d}}$ was an arbitrary perturbation with $\|\overline{\Delta \mathfrak{d}}\| \leq \hat{\rho}$ we have

$$\rho(\bar{\mathfrak{d}}) \leq \hat{\rho} = \rho(\mathfrak{d}(\mathcal{T})),$$

proving the Lemma. □