A New Complexity Result on Solving the Markov Decision Problem

Yinyu Ye *

September 2002, Revised October 1 2004

Abstract

We present a new complexity result on solving the Markov decision problem (MDP) with n states and a number of actions for each state, a special class of real-number linear programs with the Leontief matrix structure. We prove that, when the discount factor θ is strictly less than 1, the problem can be solved in at most $O(n^{1.5}(\log \frac{1}{1-\theta} + \log n))$ classical interior-point method iterations and $O(n^4(\log \frac{1}{1-\theta} + \log n))$ arithmetic operations. Our method is a *combinatorial* interior-point method related to the work of Ye [30] and Vavasis and Ye [26]. To our knowledge, this is the first *strongly polynomial-time* algorithm for solving the MDP when the discount factor is a constant less than 1.

1 Introduction

Complexity theory is arguably the foundational stone of computer algorithms. The goal of the theory is twofold: to develop criteria for measuring the effectiveness of various algorithms (and thus, to be able to compare algorithms using these criteria), and to asses the inherent difficulty of various problems.

The term "complexity" refers to the amount of resources required by a computation. We will focus on a particular resource namely, the computing time. In complexity theory, however, one is not interested on the execution time of a program implemented in a particular programming language, running on a particular computer over a particular input. There are too many contingent factors here. Instead, one would like to associate to an algorithm some more intrinsic measures of its time requirements. Roughly speaking, to do so one needs to fix:

*Department of Management Science and Engineering, Stanford University, Stanford, CA 94305; E-mail: yinyu-ye@stanford.edu. This research is supported by NSF grant DMS-0306611. The author would like to thank Kahn Mason, Ben Van Roy and Pete Veinott for many insightful discussions on this subject.

- a notion of *input size*,
- a set of *basic operations*, and
- a *cost* for each basic operation.

The last two allow one to define the *cost of a computation*.

The selection of a set of basic operations is generally easy. For the algorithms we will consider in this paper, the obvious choice is the set $\{+, -, \times, /, \leq\}$ of the four arithmetic operations and the comparison. Selecting a notion of input size and a cost for the basic operations is more delicate and depends on the kind of data dealt with by the algorithm. Some kinds can be represented within a fixed amount of computer memory, some others require a variable amount depending on the data at hand.

Examples of the first are fixed-precision floating-point numbers. Any such number is stored in a fixed amount of memory (usually 32 or 64 bits). For this kind of data the size of an element is usually taken to be 1 and consequently to have *unit size*. Examples of the second are integer numbers which require a number of bits approximately equal to the logarithm of their absolute value. This logarithm is usually referred to as the *bit size* of the integer. These ideas also apply for rational numbers.

Similar considerations apply for the cost of arithmetic operations. The cost of operating two unit-size numbers is taken to be 1 and, as expected, is called *unit cost*. In the bit-size case, the cost of operating two numbers is the product of their bit-sizes (for multiplications and divisions) or its maximum (for additions, subtractions, and comparisons).

The consideration of integer or rational data with their associated bit size and bit cost for the arithmetic operations is usually referred to as the *Turing model of computation* (e.g. see [22]). The consideration of idealized reals with unit size and unit cost is today referred as the *BSS model of computation* (from Blum, Shub and Smale [2]), which is what we have used in this paper, with exact real arithmetic operations (i.e., ignoring round-off errors).

A basic concept related to both the Turing and the BSS models of computation is that of *polynomial time*. An algorithm $\mathcal{A}$ is said to be a polynomial time algorithm if the running time of all instances of the problem is bounded above by a polynomial in the size of the problem. A problem can be solved in polynomial time if there is a polynomial time algorithm solving the problem.

The notion of polynomial time is usually taken as the formal counterpart of the more informal notion of efficiency. One not fully identify polynomial-time with efficiency since high degree polynomial bounds can hardly mean efficiency. Yet, many basic problems admitting polynomial-time algorithms can actually be efficiently solved. On the other hand, the notion of polynomial-time is robust and it is the ground upon which complexity theory is built.

Linear programming (LP) has played a distinguished role in complexity theory. It has a primal-dual form:

$$\begin{array}{rl} \text{Primal: minimize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}, \end{array} \tag{1}$$

and

$$\begin{array}{rl} \text{Dual: maximize} & \mathbf{b}^T\mathbf{y} \\ \text{subject to} & \mathbf{s} = \mathbf{c} - A^T\mathbf{y} \geq \mathbf{0}, \end{array} \tag{2}$$

where $A \in \mathbb{R}^{m\times n}$ is a given real matrix with rank m, $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$ are given real vectors, and $\mathbf{x} \in \mathbb{R}^n$ and $(\mathbf{y} \in \mathbb{R}^m, \mathbf{s} \in \mathbb{R}^n)$ are unknown real vectors. Vector $\mathbf{s}$ is often called dual slack vector. We denote the LP problem as $LP(A, \mathbf{b}, \mathbf{c})$.

In one sense LP is a continuous optimization problem since the goal is to minimize a linear objective function over a convex polyhedron. But it is also a combinatorial problem involving selecting an extreme point among a finite set of possible vertices. An optimal solution of a linear program always lies at a vertex of the feasible polyhedron. Unfortunately, the number of vertices associated with a set of n inequalities in m variables can be exponential in the dimensions—in this case, up to $n!/m!(n-m)!$. Except for small values of m and n, this number is sufficiently large to prevent examining all possible vertices for searching an optimal vertex.

The LP problem is polynomial solvable under the Turing model of computation, proved by Khachiyan [11] and also by Karmarkar [10] and many others. But the problem, whether there is a polynomial-time algorithm for LP under the BSS model of computation, remains open. It turns out that two instances of the LP problem with the same (unit) size may result in drastically different performances under an interior-point algorithm. This has lead during the past years to some research developments relating complexity of interior-point algorithms with certain "condition" measures for linear programming ([24, 29, 7, 9]). One particular example is the Vavasis-Ye algorithm, which interleaves small steps with longer *layered least-squares* (LLS) steps to follow the central path, see [26, 17, 20, 21]. The algorithm, which will be called "layered-step interior point" (LIP), terminates in a finite number of steps. Furthermore, the total number of iterations depends only on A: the running time is $O(n^{3.5}c(A))$ iterations, where $c(A)$ is the condition measure of the full-rank constraint matrix A defined in [26]:

$$c(A) = O(\log(\bar{\chi}_A) + \log n), \tag{3}$$

where

$$\bar{\chi}_A = \max\{\|(A_B)^{-1}A\| : \quad A_B \in \mathbb{R}^{m\times m} \text{ is a basic matrix of } A\}$$

This is in contrast to other interior point methods, whose complexity depend on the vectors $\mathbf{b}$ and $\mathbf{c}$ as well as on the matrix A. This is important because there are many classes of problems in which A is "well-behaved" but $\mathbf{b}$ and $\mathbf{c}$ are arbitrary vectors. $\|\cdot\|$, without subscript, incidates 2-norm through out this paper.

2 The Markov Decision Problem

In this paper, we present a new complexity result for solving the real-number Markov decision problem (MDP), a special class of real-number linear programs with the Leontief matrix structure due to de Ghellinck [5], D'Epenoux [6] and Manne [14] (see also the recent survey by Van Roy [25]):

$$\begin{array}{rccccc} \text{minimize} & \mathbf{c}_1^T\mathbf{x}_1 & \ldots + \mathbf{c}_j^T\mathbf{x}_j + \ldots & +\mathbf{c}_n^T\mathbf{x}_n & & \\ \text{subject to} & (E_1-\theta P_1)\mathbf{x}_1 & \ldots + (E_j-\theta P_j)\mathbf{x}_j + \ldots & (E_n-\theta P_n)\mathbf{x}_n & = & \mathbf{e}, \\ & \mathbf{x}_1, & \ldots \quad \mathbf{x}_j, \quad \ldots & \mathbf{x}_n, & \geq & \mathbf{0}, \end{array}$$

where $\mathbf{e}$ is the vector of all ones, E_j is the $n \times k$ matrix whose jth row are all ones and everywhere else are zeros, P_j is an $n \times k$ Markov or column stochastic matrix such that

$$\mathbf{e}^T P_j = \mathbf{e}^T \quad \text{and} \quad P_j \geq \mathbf{0}, \quad j = 1, \ldots, n.$$

Here, decision vector of $\mathbf{x}_j \in \mathbb{R}^k$ represents the decision variables associated with jth state's k actions and $\mathbf{c}_j$ is its cost vector corresponding to the k action variables. The optimal solution to the problem will select one optimal action from every state, which form an optimal feasible basis. The dual of the problem is given by

$$\begin{array}{rccc} \text{maximize} & \mathbf{e}^T\mathbf{y} & & \\ \text{subject to} & (E_1-\theta P_1)^T\mathbf{y} & \leq & \mathbf{c}_1, \\ & \ldots & \ldots & \ldots \\ & (E_j-\theta P_j)^T\mathbf{y} & \leq & \mathbf{c}_j, \\ & \ldots & \ldots & \ldots \\ & (E_n-\theta P_n)^T\mathbf{y} & \leq & \mathbf{c}_n. \end{array}$$

If we sort the decision variables by actions, the MDP can be also written as:

$$\begin{array}{rccccc} \text{minimize} & (\mathbf{c}^1)^T\mathbf{x}^1 & \ldots + (\mathbf{c}^i)^T\mathbf{x}^i + \ldots & +(\mathbf{c}^k)^T\mathbf{x}^k & & \\ \text{subject to} & (I-\theta P^1)\mathbf{x}^1 & \ldots + (I-\theta P^i)\mathbf{x}^i + \ldots & +(I-\theta P^k)\mathbf{x}^k & = & \mathbf{e}, \\ & \mathbf{x}^1, & \ldots \quad \mathbf{x}^i, \quad \ldots & \mathbf{x}^k, & \geq & \mathbf{0}. \end{array} \tag{4}$$

And its dual (by adding slack variables) is

$$\begin{array}{rccc} \text{maximize} & \mathbf{e}^T\mathbf{y} & & \\ \text{subject to} & (I-\theta P^1)^T\mathbf{y} + \mathbf{s}^1 & = & \mathbf{c}^1, \\ & \ldots & \ldots & \ldots \\ & (I-\theta P^i)^T\mathbf{y} + \mathbf{s}^i & = & \mathbf{c}^i, \\ & \ldots & \ldots & \ldots \\ & (I-\theta P^k)^T\mathbf{y} + \mathbf{s}^k & = & \mathbf{c}^k, \\ & \mathbf{s}^1, \ldots, \mathbf{s}^i, \ldots, \mathbf{s}^k & \geq & \mathbf{0}. \end{array} \tag{5}$$

Here, $\mathbf{x}^i \in \mathbb{R}^n$ represents the decision variables of all states for action i, I is the $n \times n$ identity matrix, and P^i, $i = 1, \dots, k$, is an $n \times n$ Markov matrix ($\mathbf{e}^T P^i = \mathbf{e}^T$ and $P^i \geq 0$). Comparing to the LP standard form, we have

$$A = [I - \theta P^1, \dots, I - \theta P^k] \in \mathbb{R}^{n \times nk}, \quad \mathbf{b} = \mathbf{e} \in \mathbb{R}^n, \quad \text{and} \quad \mathbf{c} = (\mathbf{c}^1; \dots; \mathbf{c}^k) \in \mathbb{R}^{nk}.$$

In the MDP, θ is the so called discount factor such that

$$\theta = \frac{1}{1+r} \leq 1,$$

where r is the interest rate and it is assumed strictly positive in this writing so that $0 \leq \theta < 1$. The problem is to find the best action for each state so that the total cost is minimized.

There are four current effective methods for solving the MDP: the value iteration method, the policy iteration method, regular LP interior-point algorithms, and the Vavasis-Ye algorithm. In terms of the worst-case complexity bound on the number of arithmetic operations, they (without a constant factor) are summarized in the following table for $k = 2$ (see Littman et al. [13], Mansour and Singh [15] and references therein).

Value-Iteration	Policy-Iteration	LP-Algorithms	Vavasis-Ye	New Method
$n^2 c(P_j, \mathbf{c}_j, \theta) \cdot \frac{1}{1-\theta}$	$n^3 \cdot \frac{2^n}{n}$	$n^3 c(P_j, \mathbf{c}_j, \theta)$	$n^6 c'(P_j, \theta)$	$n^4 \cdot \log \frac{1}{1-\theta}$

where $c(P_j, \mathbf{c}_j, \theta)$ and $c'(P_j, \theta)$ are conditional measures of data $(P_j, \mathbf{c}_j, \theta)$, $j = 1, \dots, k$. When data $(P_j, \mathbf{c}_j, \theta)$ are rational numbers, these condition measures are generally bounded by the total bit size of the input.

In this paper we develop a "combinatorial" interior-point algorithm, when the discount factor θ is strictly less than 1, to solve the MDP problem in at most $O(n^{1.5}(\log \frac{1}{1-\theta} + \log n))$ interior-point method iterations and total $O(n^4(\log \frac{1}{1-\theta} + \log n))$ arithmetic operations. Note that this bound is only dependent logarithmically on $1 - \theta$. If θ is rational, then $\log \frac{1}{1-\theta}$ is bounded above by the bit length of θ. This is because, when $\theta = q/p$ where p and q are positive integers with $p \geq q + 1$, we have

$$\log \frac{1}{1-\theta} = \log(p) - \log(p - q) \leq \log(p) \leq \text{bit length of } \theta.$$

Our method is closely related to the work of Ye [30] and Vavasis and Ye [26]. To our knowledge, this is the first *strongly polynomial-time* algorithm for solving the MDP when the discount factor (interest rate) is a constant less than 1 (greater than 0).

3 LP Theorems and Interior-Point Algorithms

We first describe few general LP theorems and a classical predictor-corrector interior-point algorithm.

3.1 Optimality condition and central path

The *optimality conditions* for all optimal solution pairs of $LP(A, \mathbf{b}, \mathbf{c})$ may be written as follows:

$$
\begin{aligned}
A\mathbf{x} &= \mathbf{b}, \\
A^T\mathbf{y} + \mathbf{s} &= \mathbf{c}, \\
SX\mathbf{e} &= \mathbf{0}, \\
\mathbf{x} \geq \mathbf{0}, &\quad \mathbf{s} \geq \mathbf{0}
\end{aligned}
\tag{6}
$$

where X denotes diag($\mathbf{x}$) and S denotes diag($\mathbf{s}$), and $\mathbf{0}$ and $\mathbf{e}$ denotes the vector of all 0's and all 1's, respectively. The third equation is often referred as the complementarity condition.

If $LP(A, \mathbf{b}, \mathbf{c})$ has an optimal solution pair, then there exists a unique index set $B^* \subset \{1, \ldots, n\}$ and $N^* = \{1, \ldots, n\} \setminus B^*$, such that every $\mathbf{x}$, satisfying

$$A_{B^*}\mathbf{x}_{B^*} = \mathbf{b}, \quad \mathbf{x}_{B^*} \geq \mathbf{0}, \quad \mathbf{x}_{N^*} = \mathbf{0},$$

is an optimal solution for the primal; and every $(\mathbf{y}, \mathbf{s})$, satisfying

$$\mathbf{s}_{B^*} = \mathbf{c}_{B^*} - A_{B^*}^T\mathbf{y} = \mathbf{0}, \quad \mathbf{s}_{N^*} = \mathbf{c}_{N^*} - A_{N^*}^T\mathbf{y} \geq \mathbf{0},$$

is an optimal solution for the dual. This partition is called the strict complementarity partition, since a "strictly" complementary pair $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ exists, meaning

$$A_{B^*}\mathbf{x}_{B^*} = \mathbf{b}, \quad \mathbf{x}_{B^*} > \mathbf{0}, \quad \mathbf{x}_{N^*} = \mathbf{0},$$

and

$$\mathbf{s}_{B^*} = \mathbf{c}_{B^*} - A_{B^*}^T\mathbf{y} = \mathbf{0}, \quad \mathbf{s}_{N^*} = \mathbf{c}_{N^*} - A_{N^*}^T\mathbf{y} > \mathbf{0}.$$

Here, for example, subvector $\mathbf{x}_B$ contains all x_i for $i \in B \subset \{1, \ldots, n\}$.

Consider the following equations:

$$
\begin{aligned}
A\mathbf{x} &= \mathbf{b}, \\
A^T\mathbf{y} + \mathbf{s} &= \mathbf{c}, \\
SX\mathbf{e} &= \mu\mathbf{e}, \\
\mathbf{x} > \mathbf{0}, &\quad \mathbf{s} > \mathbf{0}.
\end{aligned}
\tag{7}
$$

These equations always have a unique solution for any $\mu > 0$ provided that the primal and dual problem both have interior feasible points ($\mathbf{x} > \mathbf{0}, \mathbf{s} > \mathbf{0}$). The solution to these equations, written $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$, is called the *central path point* for μ, and the aggregate of all points, as μ ranges from 0 to ∞, is the *central path* of the LP problem, see [1, 16]. Note that as $\mu \to 0^+$, central path equation (7) approaches optimality condition (6), and $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ approaches an optimal solution.

The following is a geometric property of the central path:

Lemma 1 *Let* $(\mathbf{x}(\mu), \mathbf{y}(\mu), \mathbf{s}(\mu))$ *and* $(\mathbf{x}(\mu'), \mathbf{y}(\mu'), \mathbf{s}(\mu'))$ *be two central path points such that* $0 \le \mu' < \mu$. *Then for any* i,

$$s(\mu')_i \le ns(\mu)_i \quad \textit{and} \quad x(\mu')_i \le nx(\mu)_i.$$

In particular, given any optimal $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$, *we have, for any* $\mu > 0$ *and any* i,

$$s_i^* \le ns(\mu)_i \quad \textit{and} \quad x_i^* \le nx(\mu)_i.$$

Proof. The lemma was proved in [26]. We reprove the second part here to give the reader an insight view on the structure. Since $\mathbf{x}(\mu) - \mathbf{x}^* \in \mathcal{N}(A)$ and $\mathbf{s}(\mu) - \mathbf{s}^* \in \mathcal{R}(A^T)$ we have $(\mathbf{x}(\mu) - \mathbf{x}^*)^T(\mathbf{s}(\mu) - \mathbf{s}^*) = 0$. Since $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ is optimal, we have $(\mathbf{x}^*)^T\mathbf{s}^* = 0$. These two equations imply

$$(\mathbf{x}^*)^T\mathbf{s}(\mu) + (\mathbf{s}^*)^T\mathbf{x}(\mu) = \mathbf{x}(\mu)^T\mathbf{s}(\mu) = n\mu.$$

Dividing μ on both sides and noting $x(\mu)_i s(\mu)_i = \mu$ for each $i = 1, \dots, n$, we have

$$\sum_{i=1}^n \frac{x_i^*}{x(\mu)_i} + \sum_{i=1}^n \frac{s_i^*}{s(\mu)_i} = n.$$

Because each term in the summation is nonnegative, for any $i = 1, \dots, n$ we have

$$\frac{x_i^*}{x(\mu)_i} \le n \quad \text{and} \quad \frac{s_i^*}{s(\mu)_i} \le n.$$

∎

3.2 Predictor-corrector interior-point algorithm

In the predictor-corrector path-following interior point method of Mizuno et al. [18], one solves (7) approximately to obtain an "approximately centered" point $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu)$ such that

$$\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) := \|SX\mathbf{e}/\mu - \mathbf{e}\| \le \eta_0, \tag{8}$$

where, say, $\eta_0 = 0.2$ throughout this paper. Then each iteration decreases μ towards zero, and for each new value of μ, the approximate point $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu)$ to central path equation (7) is updated.

For any point $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ and μ where $\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) \leq 1/5$, it has been proved that

$$\begin{array}{c} (3/4)s_i \leq s(\mu)_i \leq (5/4)s_i \\ (3/4)x_i \leq x(\mu)_i \leq (5/4)x_i. \end{array} \tag{9}$$

In general, since each diagonal entry of SX is between $\mu(1-\eta_0)$ and $\mu(1+\eta_0)$, we have the two inequalities

$$\begin{array}{c} \sqrt{\mu(1-\eta_0)} \leq \|(SX)^{1/2}\| \leq \sqrt{\mu(1+\eta_0)} \\ \frac{1}{\sqrt{\mu(1+\eta_0)}} \leq \|(SX)^{-1/2}\| \leq \frac{1}{\sqrt{\mu(1-\eta_0)}}, \end{array} \tag{10}$$

which will be used frequently during the upcoming analysis.

Given a near central path point $(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu)$ such that $\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) \leq \eta_0$, the newly decreased μ can be calculated by solving two related least squares problems in a predictor step. Let

$$D = X^{-1}S, \tag{11}$$

and let $(\delta\bar{\mathbf{y}}, \delta\bar{\mathbf{s}})$ be the solution to a weighted least-squares problem

$$\min \|D^{-1/2}(\delta\mathbf{s} + \mathbf{s})\| \text{ subject to } \delta\mathbf{s} = -A^T\delta\mathbf{y} \text{ or } \delta\mathbf{s} \in \mathcal{R}(A^T),$$

where $\mathcal{R}(A^T)$ denotes the range space of A^T; and let $\delta\bar{\mathbf{x}}$ be the solution to the related weighted least-squares problem

$$\min \|D^{1/2}(\delta\mathbf{x} + \mathbf{x})\| \text{ subject to } A\delta\mathbf{x} = \mathbf{0} \text{ or } \delta\mathbf{x} \in \mathcal{N}(A),$$

where $\mathcal{N}(A)$ denotes the null space of A.

It is not hard to see that $(\delta\bar{\mathbf{x}}, \delta\bar{\mathbf{y}}, \delta\bar{\mathbf{s}})$ satisfy the following equations:

$$\begin{array}{r} A\delta\bar{\mathbf{x}} = \mathbf{0}, \\ A^T\delta\bar{\mathbf{y}} + \delta\bar{\mathbf{s}} = \mathbf{0}, \\ D^{1/2}\delta\bar{\mathbf{x}} + D^{-1/2}\delta\bar{\mathbf{s}} = -X^{1/2}S^{1/2}\mathbf{e}. \end{array} \tag{12}$$

These equations imply that

$$\|D^{1/2}\delta\bar{\mathbf{x}}\|^2 + \|D^{-1/2}\delta\bar{\mathbf{s}}\|^2 = n\mu. \tag{13}$$

The new iterate point $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})$, defined by

$$(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}) = (\mathbf{x}, \mathbf{y}, \mathbf{s}) + \alpha(\delta\bar{\mathbf{x}}, \delta\bar{\mathbf{y}}, \delta\bar{\mathbf{s}})$$

for a suitable step size $\alpha \in (0, 1]$, is a strictly feasible point. Furthermore,

$$\eta(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}, \mu(1-\alpha)) \leq 2\eta_0.$$

Then, one takes a corrector step—uses the Newton method starting from $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})$ to generate a new $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ such that $\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu(1-\alpha)) \leq \eta_0$. Thus, the new μ is reduced by a factor $(1-\alpha)$ from μ. One can further show that the step size can be greater than $\frac{1}{4\sqrt{n}}$ in every iteration. Thus, the predictor-corrector interior-point algorithm reduces μ to μ' $(< \mu)$ in at most $O(\sqrt{n}\log(\mu/\mu'))$ iterations, while maintaining suitable proximity to the central path. Moreover, the average arithmetic operations per iteration is $O(n^{2.5})$. This eventually results in a polynomial time algorithm in the bit-size computation model—$O(n^3 L)$, see also [23, 8, 12, 19].

3.3 Properties of the Markov Decision Problem

For simplicity, we fix k, the number of actions taken by each state, to 2 in the rest of the paper. Then, the MDP can be rewritten as:

$$\begin{array}{lccl} \text{minimize} & (\mathbf{c}^1)^T\mathbf{x}_1 & +(\mathbf{c}^2)^T\mathbf{x}_2 & \\ \text{subject to} & (I-\theta P^1)\mathbf{x}^1 & +(I-\theta P^2)\mathbf{x}^2 & = \mathbf{e}, \\ & \mathbf{x}^1, & \mathbf{x}^2 & \geq \mathbf{0}. \end{array} \tag{14}$$

And its dual is

$$\begin{array}{lcl} \text{maximize} & \mathbf{e}^T\mathbf{y} & \\ \text{subject to} & (I-\theta P^1)^T\mathbf{y} + \mathbf{s}^1 & = \mathbf{c}^1, \\ & (I-\theta P^2)^T\mathbf{y} + \mathbf{s}^2 & = \mathbf{c}^2, \\ & \mathbf{s}^1, \mathbf{s}^2 & \geq \mathbf{0}. \end{array} \tag{15}$$

Comparing to the LP standard form, we have

$$A = [I-\theta P^1,\ I-\theta P^2] \in \mathbb{R}^{n\times 2n}, \quad \mathbf{b} = \mathbf{e} \in \mathbb{R}^n, \quad \text{and} \quad \mathbf{c} = (\mathbf{c}^1; \mathbf{c}^2) \in \mathbb{R}^{2n}.$$

Note that any feasible basis A_B of the MDP has the Leontief form

$$A_B = I - \theta P$$

where P is an $n \times n$ Markov matrix chosen from columns of $[P^1,\ P^2]$, and the reverse is also true. Most proofs of the following lemma can be found in Dantzig [3, 4] and Veinott [27].

Lemma 2 *The MDP has the following properties:*

1. *Both the primal and dual MDPs have interior feasible points if* $0 \leq \theta < 1$.

2. *The feasible set of the primal MDP is bounded. More precisely,*

$$\mathbf{e}^T\mathbf{x} = \frac{n}{1-\theta},$$

where $\mathbf{x} = (\mathbf{x}^1; \mathbf{x}^2)$.

3. *Let* $\hat{\mathbf{x}}$ *be a basic feasible solution of the MDP. Then, any basic variable, say* $\hat{x}_i$, *has its value*

$$\hat{x}_i \geq 1.$$

4. *Let* B^* *and* N^* *be the optimal partition for the MDP. Then,* B^* *contains at least one feasible basis, i.e.,* $|B^*| \geq n$ *and* $|N^*| \leq n$*; and for any* $j \in B^*$ *there is an optimal solution* $\mathbf{x}^*$ *such that*

$$x_j^* \geq 1.$$

5. *Let* A_B *be any feasible basis and* A_N *be any submatrix of the rest columns of the MDP constraint matrix, then*

$$\|(A_B)^{-1}A_N\| \leq \frac{2n\sqrt{n}}{1-\theta}.$$

PROOF. The proof of this lemma is straightforward and based on the expressions

$$\mathbf{e}^T P = \mathbf{e}^T \quad \text{and} \quad (I - \theta P)^{-1} = I + \theta P + \theta^2 P^2 + \ldots,$$

and the fact that P^k remains a Markov matrix for $k = 1, 2, \ldots$.

It is easy to verify that

$$\mathbf{x}_1 = (I - \theta P^1)^{-1}\mathbf{e}/2 \quad \text{and} \quad \mathbf{x}_2 = (I - \theta P^2)^{-1}\mathbf{e}/2$$

is an intetior feasile solution to the primal; and

$$\mathbf{y} = -\gamma\mathbf{e}$$

is an interior feasible solution to the dual for sufficiently large γ. This proves (1).

Left multiplying $\mathbf{e}^T$ to both sides of the constraints of the primal, we have

$$\mathbf{e}^T A\mathbf{x} = (1-\theta)\mathbf{e}^T\mathbf{x} = n$$

which proves (2).

Since any basic feasible solution has the form

$$(I - \theta P)^{-1}\mathbf{e} = (I + \theta P + \theta^2 P^2 + \ldots)\mathbf{e} \geq \mathbf{e},$$

(3) is proved.

Since the MDP has at least one basic optimal solution, the proof of (4) follows the statement of (3).

The proof of (5) is more demanding. Note that any feasible basis

$$A_B^{-1} = (I - \theta P)^{-1} = I + \theta P + \theta^2 P^2 + \dots$$

where P is a Markov matrix. Let the ith column of P be P_i. Then we have

$$\|P_i\| \leq \mathbf{e}^T P_i = 1$$

because $\mathbf{e}^T P_i$ is the 1-norm of $P_i \geq \mathbf{0}$ and the 2-norm is less than or equal to the 1-norm. Now, for any vector $\mathbf{a} \in \mathbb{R}^n$, we have

$$\|P\mathbf{a}\| = \|\sum_{i=1}^{n} P_i \cdot a_i\| \leq \sum_{i=1}^{n} \|P_i\| \cdot |a_i| \leq \sum_{i=1}^{n} |a_i| = \|\mathbf{a}\|_1$$

and

$$\|\mathbf{a}\|_1 \leq \sqrt{n} \cdot \|\mathbf{a}\|,$$

which imply that

$$\|P\mathbf{a}\| \leq \sqrt{n} \cdot \|\mathbf{a}\|$$

or

$$\|P\| \leq \sqrt{n}$$

for any Markov matrix P.

Furthermore, each component of $A_N \in \mathbb{R}^{n \times t}$ ($t \leq 2n$) is between -1 and 1. Let the ith column of A_N be $(A_N)_i$. Then, $\|(A_N)_i\| \leq \sqrt{n}$ and, for any vector $\mathbf{d} \in \mathbb{R}^t$, we have

$$\|A_N \mathbf{d}\| \leq \sum_{i=1}^{t} \|(A_N)_i\| \cdot |d_i| \leq \sqrt{n} \cdot \|\mathbf{d}\|_1 \leq \sqrt{n} \cdot \sqrt{t} \cdot \|\mathbf{d}\| \leq 2n \cdot \|\mathbf{d}\|.$$

Finaly, for any vector $\mathbf{d} \in \mathbb{R}^t$,

$$\begin{aligned}
\|A_B^{-1} A_N \mathbf{d}\| &= \|(I + \theta P + \theta^2 P^2 + \dots) A_N \mathbf{d}\| \\
&\leq \|A_N \mathbf{d}\| + \theta \|P A_N \mathbf{d}\| + \theta^2 \|P^2 A_N \mathbf{d}\| + \dots \\
&\leq \|A_N \mathbf{d}\| + \theta \|P\| \cdot \|A_N \mathbf{d}\| + \theta^2 \|P^2\| \cdot \|A_N \mathbf{d}\| + \dots \\
&\leq \|A_N \mathbf{d}\| + \theta \sqrt{n} \cdot \|A_N \mathbf{d}\| + \theta^2 \sqrt{n} \cdot \|A_N \mathbf{d}\| + \dots \\
&\leq \frac{\sqrt{n}}{1 - \theta} \cdot \|A_N \mathbf{d}\| \\
&\leq \frac{\sqrt{n}}{1 - \theta} \cdot 2n \cdot \|\mathbf{d}\|,
\end{aligned}$$

or

$$\|(A_B)^{-1} A_N\| \leq \frac{2n\sqrt{n}}{1 - \theta}.$$

∎

Using the lemma, we may safely assume that $\mathbf{c} \geq \mathbf{0}$ since using $\mathbf{c} + \gamma\mathbf{e}$ for any number γ is equivalent to using $\mathbf{c}$ in the MDP.

We also remark that the condition measure $\bar{\chi}_A$ mentioned earlier and used in [26, 17, 20] cannot be bounded by $\frac{2n\sqrt{n}}{1-\theta}$ of (5) of the lemma, since it is the maximum over all possible (feasible or infeasible) bases. Consider a two-state and two-action MDP where

$$A = \begin{bmatrix} 1-\theta & 0 & 1-\theta(1-\epsilon) & 0 \\ 0 & 1-\theta & -\theta \cdot \epsilon & 1-\theta \end{bmatrix}$$

Here, for any given $\theta > 0$, $\|(A_B)^{-1}A\|$ can be arbitrarily large as $\epsilon \to 0^+$ when

$$A_B = \begin{pmatrix} 1-\theta & 1-\theta(1-\epsilon) \\ 0 & -\theta \cdot \epsilon \end{pmatrix}.$$

In fact, all other condition measures ([24, 29, 7, 9]) used in complexity analyses for general LP can be arbitrarily bad for the MDP. In other words, our new complexity result achieved in this paper cannot be derived by simply proving that the MDP is well conditioned based on one of these measures.

4 Combinatorial interior-point method for solving the MDP

We now develop a combinatorial interior-point method for solving the MDP. The method has at most n major steps, where each major step identifies and eliminates at least one variable in N^* and then the method proceeds to solve the MDP with at least one variable less. Each major step uses at most $O(n^{0.5}(\log \frac{1}{1-\theta} + \log n))$ predictor-corrector iterations. This elimination process shares the same spirit of the "build-down" scheme analyzed in [30]

4.1 Complexity to compute a near central-path point

Like any other interior-point algorithm, our method needs a near central-path point pair satisfying (8) to start with. We now analyze the complexity to generate such a point, and show this can be accomplished in $O(n \log \frac{1}{1-\theta})$ classical interior-point algorithm iterations.

Let

$$(\mathbf{x}^i)^0 = (I - \theta P^i)^{-1} e, \quad i = 1, 2$$

and

$$\mathbf{x}^0 = \begin{pmatrix} \frac{1}{2}(\mathbf{x}^1)^0 \\ \frac{1}{2}(\mathbf{x}^2)^0 \end{pmatrix}.$$

Thus, $\mathbf{x}^0$ is an interior feasible point for the MDP and

$$\mathbf{x}^0 \geq \frac{1}{2}\mathbf{e} \in \mathbb{R}^{2n}.$$

Let

$$\mathbf{y}^0 = -\gamma\mathbf{e} \quad \text{and} \quad \mathbf{s}^0 = \begin{pmatrix} (\mathbf{s}^1)^0 \\ (\mathbf{s}^2)^0 \end{pmatrix} = \begin{pmatrix} \mathbf{c}^1 + \gamma(1-\theta)\mathbf{e} \\ \mathbf{c}^2 + \gamma(1-\theta)\mathbf{e}. \end{pmatrix}$$

where γ is chosen sufficiently large such that

$$\mathbf{s}^0 > 0 \quad \text{and} \quad \gamma \geq \frac{\mathbf{c}^T\mathbf{x}^0}{n}.$$

Denote $\mu^0 = (\mathbf{x}^0)^T\mathbf{s}^0/2n$ and consider the potential function

$$\phi(\mathbf{x}, \mathbf{s}) = 2n\log(\mathbf{s}^T\mathbf{x}) - \sum_{j=1}^{2n}\log(s_jx_j).$$

Let $c_j(\geq 0)$ be the jth coefficient of $\mathbf{c} = (\mathbf{c}^1; \mathbf{c}^2) \geq \mathbf{0}$. Then, we have

$$\begin{aligned}
\phi(\mathbf{x}^0, \mathbf{s}^0) &= 2n\log(\mathbf{c}^T\mathbf{x}^0 + \gamma(1-\theta)\frac{n}{1-\theta}) - \sum_{j=1}^{2n}\log(s_j^0x_j^0) \\
&\leq 2n\log(\mathbf{c}^T\mathbf{x}^0 + \gamma\cdot n) - \sum_{j=1}^{2n}\log(s_j^0/2) \quad (\text{since } x_j^0 \geq 1/2) \\
&= 2n\log(2n) - \sum_{j=1}^{2n}\log\frac{2n(c_j/2 + \gamma(1-\theta)/2)}{\mathbf{c}^T\mathbf{x}^0 + \gamma\cdot n} \\
&\leq 2n\log(2n) - \sum_{j=1}^{2n}\log\frac{n\gamma(1-\theta)}{\mathbf{c}^T\mathbf{x}^0 + \gamma\cdot n} \quad (\text{since } \mathbf{c} \geq \mathbf{0}) \\
&\leq 2n\log(2n) - \sum_{j=1}^{2n}\log\frac{n\gamma(1-\theta)}{2\gamma\cdot n} \quad (\text{since } \gamma \geq \mathbf{c}^T\mathbf{x}^0/n) \\
&= 2n\log(2n) + 2n\log(\frac{2}{1-\theta}).
\end{aligned}$$

Therefore, using the primal-dual potential reduction algorithm, we can generate an approximate central path point $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{s}^0)$ such that

$$\eta(\mathbf{x}^0, \mathbf{y}^0, \mathbf{s}^0, \mu^0) \leq \eta_0.$$

in at most $O(n(\log\frac{2}{1-\theta} + \log n))$ interior-point algorithm iterations where each iteration uses $O(n^3)$ arithmetic operations, see [28],

4.2 Eliminating variables in N^*

Now we present the major step of the method to identify and eliminate variables in N^*. We start with the following lemma for a central path property of the MDP:

Lemma 3 *For any $\mu \in (0, \mu^0]$, the central path pair of (14) and (15) satisfies*

$$x(\mu)_j \leq \frac{n}{1-\theta} \quad \textit{and} \quad s(\mu)_j \geq \frac{1-\theta}{n}\mu \quad \textit{for every } j = 1, \dots, 2n;$$

and

$$x(\mu)_j \geq \frac{1}{2n} \quad \textit{and} \quad s(\mu)_j \leq 2n\mu \quad \textit{for evey } j \in B^*.$$

PROOF. The upper bound on $x(\mu)_j$ is from that the sum of all primal variables is equal to $\frac{n}{1-\theta}$. The lower bounds on $x(\mu)_j$ are direct results of Lemmas 1 and 2(4) and noting there are $2n$ variables in (14). The bounds on $s(\mu)_j$ are from the central path equations of (7). ∎

Define a gap factor

$$g = \frac{10n^2(1+\eta_0)}{(1-\theta)\sqrt{1-\eta_0}} (> 1). \tag{16}$$

For any given approximate central path point $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ such that

$$\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) \leq \eta_0,$$

define

$$J_1(\mu) = \{j : \ s_j \leq \frac{8n\mu}{3}\},$$

$$J_3(\mu) = \{j : \ s_j \geq \frac{8n\mu \cdot g}{3}\}$$

and $J_2(\mu)$ be the rest of indices. Thus, for any $j_1 \in J_1(\mu)$ and $j_3 \in J_3(\mu)$, we have

$$\frac{s_{j_1}}{s_{j_3}} \leq \frac{1}{g} < 1. \tag{17}$$

From Lemma 3 and (9), for any $j \in B^*$, we observe

$$s_j = \frac{s_j}{s(\mu)_j} s(\mu)_j \leq \frac{s_j}{s(\mu)_j} 2n\mu \leq \frac{4}{3} 2n\mu = \frac{8n\mu}{3}.$$

Therefore, we have

Lemma 4 *Let $J_1(\mu)$ be defined above for (14) and (15) at any $0 < \mu \le \mu^0$. Then, every variable of B^* is in $J_1(\mu)$ or $B^* \subset J_1(\mu)$ for any $0 < \mu \le \mu^0$, and, thereby, $J_1(\mu)$ always contains an optimal basis. Moreover, since $B^* \subset J_1(\mu)$ and $g > 1$, we must have $J_3(\mu) \subset N^*$.*

$J_3(\mu) \subset N^*$ implies that every primal variable in $J_3(\mu)$ belongs to N^* and it has zero value at any primal optimal solution. Therefore, if $J_3(\mu)$ is not empty, we can eliminate every primal variable (and dual constraint) in $J_3(\mu)$ from further consideration. To restore the primal feasibility after elimination, we solve the least squares problem:

$$\min_{\delta \mathbf{x}_1} \quad \|D_1^{1/2}\delta\mathbf{x}_1\| \text{ subject to } \quad A_1\delta\mathbf{x}_1 = A_3\mathbf{x}_3.$$

Here, subscript . denotes the subvector or submatrix of index set $J.(\mu)$, and $D_1 = X_1^{-1}S_1$. Note that the problem is always feasible since A_1 contains B^* and B^* contains at least one optimal basis.

Then, we have

$$A_1(\mathbf{x}_1 + \delta\mathbf{x}_1) + A_2\mathbf{x}_2 = A_1\mathbf{x}_1 + A_2\mathbf{x}_2 + A_3\mathbf{x}_3 = \mathbf{b}.$$

The question is whether or not $\mathbf{x}_1 + \delta\mathbf{x}_1 > 0$. We show below that

Lemma 5 *Not only $A_1(\mathbf{x}_1 + \delta\mathbf{x}_1) + A_2\mathbf{x}_2 = \mathbf{b}$ and $(\mathbf{x}_1 + \delta\mathbf{x}_1; \mathbf{x}_2) > 0$, but also*

$$\eta((\mathbf{x}_1 + \delta\mathbf{x}_1; \mathbf{x}_2), \mathbf{y}, (\mathbf{s}_1; \mathbf{s}_2), \mu) \le 2\eta_0.$$

That is, they are a near central-path point pair for the same μ of (14) and (15) after eliminating every primal variables and dual constraints in $J_3(\mu)$.

Proof. Let A_B be an optimal basis contained by A_1 and B be the index set of the basis. Then

$$A_1D_1^{-1}A_1^T \succeq A_BD_B^{-1}A_B^T \succ \mathbf{0};$$

where $D_B = X_B^{-1}S_B$, $U \succeq V$ means that matrix $U - V$ is positive semidefinite, and $V \succ \mathbf{0}$ means that matrix V is positive definite. Note that

$$\delta\mathbf{x}_1 = D_1^{-1}A_1^T(A_1D_1^{-1}A_1^T)^{-1}A_3\mathbf{x}_3$$

so that

$$\begin{aligned}
\|D_1^{1/2}\delta\mathbf{x}_1\| &= \|D_1^{-1/2}A_1^T(A_1D_1^{-1}A_1^T)^{-1}A_3\mathbf{x}_3\| \\
&= \sqrt{(A_3\mathbf{x}_3)^T(A_1D_1^{-1}A_1^T)^{-1}A_3\mathbf{x}_3} \\
&\le \sqrt{(A_3\mathbf{x}_3)^T(A_BD_B^{-1}A_B^T)^{-1}A_3\mathbf{x}_3}
\end{aligned}$$

$$
\begin{aligned}
&= \|D_B^{-1/2} A_B^T (A_B D_B^{-1} A_B^T)^{-1} A_3 \mathbf{x}_3\| \\
&= \|D_B^{1/2} A_B^{-1} A_3 \mathbf{x}_3\| \\
&\le \|D_B^{1/2}\| \cdot \|A_B^{-1} A_3\| \cdot \|\mathbf{x}_3\| \\
&\le \|D_1^{1/2}\| \cdot \|A_B^{-1} A_3\| \cdot \|\mathbf{x}_3\| \quad (\text{since } B \subset J_1(\mu)) \\
&= \|S_1 (X_1 S_1)^{-1/2}\| \cdot \|A_B^{-1} A_3\| \cdot \|S_3 X_3 S_3^{-1} \mathbf{e}\| \\
&\le \|S_1\| \cdot \|(X_1 S_1)^{-1/2}\| \cdot \|A_B^{-1} A_3\| \cdot \|S_3 X_3\| \|S_3^{-1} \mathbf{e}\| \\
&\le \|S_1\| \cdot \frac{1}{\sqrt{\mu(1-\eta_0)}} \cdot \|A_B^{-1} A_3\| \cdot \mu(1+\eta_0) \|S_3^{-1}\| \sqrt{n} \\
&= \frac{\sqrt{n\mu}(1+\eta_0)}{\sqrt{1-\eta_0}} \cdot \|S_1\| \|S_3^{-1}\| \cdot \|A_B^{-1} A_3\| \\
&\le \frac{\sqrt{n\mu}(1+\eta_0)}{\sqrt{1-\eta_0}} \cdot \frac{1}{g} \cdot \frac{2n\sqrt{n}}{1-\theta} \quad (\text{from (17) and Lemma 2(5)}) \\
&= \sqrt{n\mu} \cdot \frac{1}{5\sqrt{n}} \\
&= \frac{\sqrt{\mu}}{5}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\|X_1^{-1} \delta \mathbf{x}_1\| &= \|(X_1 S_1)^{-1/2} D_1^{1/2} \delta \mathbf{x}_1\| \\
&\le \|(X_1 S_1)^{-1/2}\| \|D_1^{1/2} \delta \mathbf{x}_1\| \\
&\le \|(X_1 S_1)^{-1/2}\| \cdot \frac{\sqrt{\mu}}{5} \\
&\le \frac{1}{\sqrt{\mu(1-\eta_0)}} \cdot \frac{\sqrt{\mu}}{5} \\
&= \frac{1}{\sqrt{1-\eta_0}} \cdot \frac{1}{5} \\
&< 1,
\end{aligned}
$$

which implies

$$
\mathbf{x}_1 + \delta \mathbf{x}_1 = X_1 (\mathbf{e} + X_1^{-1} \delta \mathbf{x}_1) > 0.
$$

Furthermore,

$$
\begin{aligned}
\left\| \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 + \delta \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} - \mu \mathbf{e} \right\| &= \left\| \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} - \mu \mathbf{e} + \begin{pmatrix} S_1 \delta \mathbf{x}_1 \\ 0 \end{pmatrix} \right\| \\
&\le \left\| \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} - \mu \mathbf{e} \right\| + \left\| \begin{pmatrix} S_1 \delta \mathbf{x}_1 \\ 0 \end{pmatrix} \right\| \\
&\le \|S \mathbf{x} - \mu \mathbf{e}\| + \|S_1 \delta \mathbf{x}_1\|
\end{aligned}
$$

$$
\begin{aligned}
&\le \eta_0\mu + \|(S_1X_1)^{1/2}D_1^{1/2}\delta\mathbf{x}_1\| \\
&\le \eta_0\mu + \|(S_1X_1)^{1/2}\| \cdot \|D_1^{1/2}\delta\mathbf{x}_1\| \\
&\le \eta_0\mu + \sqrt{\mu(1+\eta_0)} \cdot \frac{\sqrt{\mu}}{5} \\
&\le \eta_0\mu + \frac{\sqrt{1+\eta_0}}{5}\mu \\
&\le 2\eta_0\mu.
\end{aligned}
$$

∎

5 Complexity to Make $J_3(\mu) \neq \emptyset$

The question now is what to do if $J_3(\mu)$ is empty at, say, initial μ^0. Well, we directly apply the predictor-corrector method described earlier, that is, we compute the predictor step (12) at $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ where $\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu^0) \le \eta_0$.

We first take care of the trivial case that $N^* = \emptyset$.

Lemma 6 *If $N^* = \emptyset$, then the solution to (12) is*

$$\delta\bar{\mathbf{s}} = -\mathbf{s} \quad \textit{and} \quad \delta\bar{\mathbf{x}} = \mathbf{0}.$$

That is, any feasible solution to (14) is an optimal solution.

Proof. If $N^* = \emptyset$, we have $\mathbf{s} \in \mathcal{R}(A^T)$. Thus, the minimal value of

$$\min \|D^{-1/2}(\delta\mathbf{s} + \mathbf{s})\| \text{ subject to } \delta\mathbf{s} = -A^T\delta\mathbf{y} \text{ or } \delta\mathbf{s} \in \mathcal{R}(A^T)$$

is 0 or $\delta\bar{\mathbf{s}} = -\mathbf{s}$. Then, from (12) we have $\delta\bar{\mathbf{x}} = \mathbf{0}$. ∎

Therefore, we may assume that

$$\epsilon^0 := \frac{1}{\sqrt{\mu^0}}\|D^{-1/2}(\delta\bar{\mathbf{s}} + \mathbf{s})\| = \frac{1}{\sqrt{\mu^0}}\|D^{1/2}\delta\bar{\mathbf{x}}\| \tag{18}$$

is strictly greater than 0. Let

$$\bar{\alpha} = \max\left\{0, 1 - \frac{\sqrt{n}\epsilon^0}{\eta_0}\right\}. \tag{19}$$

Now take a step defined by the directions of (12); we compute a new feasible iterate

$$\bar{\mathbf{x}} = \mathbf{x} + \bar{\alpha}\delta\bar{\mathbf{x}},$$

$$\bar{\mathbf{y}} = \mathbf{y} + \bar{\alpha}\delta\bar{\mathbf{y}},$$

and

$$\bar{\mathbf{s}} = \mathbf{s} + \bar{\alpha}\delta\bar{\mathbf{s}}.$$

If $\bar{\alpha} < 1$, we have the following centrality lemma.

Lemma 7 *Let $\bar{\alpha} < 1$ in (19). Then the new iterate $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}})$ is strictly feasible. Furthermore,*

$$\eta(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{s}}, \mu^0(1-\bar{\alpha})) \le 2\eta_0.$$

Proof. The lemma is clearly true if $\bar{\alpha} = 0$. Thus, we assume that $\bar{\alpha} > 0$, that is,

$$\frac{\sqrt{n}\epsilon^0}{\eta_0} < 1 \quad \text{or} \quad \bar{\alpha} = 1 - \frac{\sqrt{n}\epsilon^0}{\eta_0} > 0.$$

Let $\alpha \in (0, \bar{\alpha}]$ and let

$$\mathbf{x}(\alpha) = \mathbf{x} + \alpha\delta\bar{\mathbf{x}},$$
$$\mathbf{y}(\alpha) = \mathbf{y} + \alpha\delta\bar{\mathbf{y}},$$

and

$$\mathbf{s}(\alpha) = \mathbf{s} + \alpha\delta\bar{\mathbf{s}}.$$

Noting that (12) implies that

$$X\delta\bar{\mathbf{s}} + S\delta\bar{\mathbf{x}} = -XS\mathbf{e},$$

we have

$$\begin{aligned}
& \mu^0(1-\alpha)\cdot\eta(\mathbf{x}(\alpha), \mathbf{y}(\alpha), \mathbf{s}(\alpha), \mu^0(1-\alpha)) \\
= {} & \|S(\alpha)X(\alpha)\mathbf{e} - \mu^0(1-\alpha)\mathbf{e}\| \\
= {} & \|(1-\alpha)SX\mathbf{e} - \mu^0(1-\alpha)\mathbf{e} + \alpha^2\Delta\bar{\mathbf{s}}\delta\bar{\mathbf{x}}\| \\
\le {} & \|(1-\alpha)SX\mathbf{e} - \mu^0(1-\alpha)\mathbf{e}\| + \|\alpha^2\Delta\bar{\mathbf{s}}\delta\bar{\mathbf{x}}\| \\
\le {} & (1-\alpha)\eta_0\mu^0 + \alpha^2\|\Delta\bar{\mathbf{s}}\delta\bar{\mathbf{x}}\| \\
\le {} & (1-\alpha)\eta_0\mu^0 + \|\Delta\bar{\mathbf{s}}\delta\bar{\mathbf{x}}\| \\
= {} & (1-\alpha)\eta_0\mu^0 + \|D^{-1/2}\Delta\bar{\mathbf{s}}D^{1/2}\delta\bar{\mathbf{x}}\| \\
\le {} & (1-\alpha)\eta_0\mu^0 + \|D^{-1/2}\Delta\bar{\mathbf{s}}\|\|D^{1/2}\delta\bar{\mathbf{x}}\| \\
\le {} & (1-\alpha)\eta_0\mu^0 + \sqrt{n\mu^0}\|D^{1/2}\delta\bar{\mathbf{x}}\| \\
= {} & (1-\alpha)\eta_0\mu^0 + \frac{\sqrt{n}\epsilon^0}{\eta_0}\eta_0\mu^0 \\
= {} & (1-\alpha)\eta_0\mu^0 + (1-\bar{\alpha})\eta_0\mu^0 \\
\le {} & 2\eta_0(1-\alpha)\mu^0.
\end{aligned}$$

Now, we argue about the feasibility of $(\mathbf{x}(\alpha), \mathbf{y}(\alpha), \mathbf{s}(\alpha))$. Observe that the proximity measure $\|S(\alpha)X(\alpha)\mathbf{e}/(\mu^0(1-\alpha)) - \mathbf{e}\|$ is a continuous function of α as α ranges over $(0, \bar{\alpha}]$. We have just proved that the proximity measure is bounded by $2\eta_0 = 0.4$ for all α in this range, and in particular, this proximity is strictly less than 1. This means no $x(\alpha)_i$ or $s(\alpha)_i$ for α in this range can be equal to 0 since $\mathbf{x}(0) = \mathbf{x} > \mathbf{0}$ and $\mathbf{s}(0) = \mathbf{s} > \mathbf{0}$. By continuity, this implies that $\mathbf{s}(\alpha) > \mathbf{0}$ and $\mathbf{x}(\alpha) > \mathbf{0}$ for all $\alpha \in (0, \bar{\alpha}]$. ∎

The next lemma identifies a variable in N^* if $\epsilon^0 \neq 0$.

Lemma 8 *If $\epsilon^0 > 0$, then there must be a variable indexed $\bar{j}$ such that $\bar{j} \in N^*$, and the central-path value*

$$s(\mu)_{\bar{j}} \geq \frac{\sqrt{1-\eta_0}(1-\theta)\mu^0}{2\sqrt{2}n^{2.5}} \cdot \epsilon^0,$$

for all $\mu \in (0, \mu^0]$.

Proof. Let $\mathbf{s}^*$ be any optimal dual slack vector for (15). Since $\mathbf{s}^* - \mathbf{s} \in \mathcal{R}(A^T)$, we must have

$$\begin{aligned}
\sqrt{\mu^0}\epsilon^0 &= \|D^{-1/2}(\delta\bar{\mathbf{s}} + \mathbf{s})\| \\
&\leq \|D^{-1/2}\mathbf{s}^*\| \\
&= \|(XS)^{-1/2}X\mathbf{s}^*\| \\
&= \|(XS)^{-1/2}\| \cdot \|X\mathbf{s}^*\| \\
&\leq \frac{1}{\sqrt{\mu^0(1-\eta_0)}} \cdot \|X\mathbf{s}^*\| \\
&\leq \frac{1}{\sqrt{\mu^0(1-\eta_0)}} \cdot \|X\|\|\mathbf{s}^*\| \\
&\leq \frac{1}{\sqrt{\mu^0(1-\eta_0)}} \cdot \frac{n}{1-\theta}\|\mathbf{s}^*\|
\end{aligned}$$

Thus,

$$\|\mathbf{s}^*\|_\infty \geq \frac{\|\mathbf{s}^*\|}{\sqrt{2n}} \geq \frac{\sqrt{1-\eta_0}(1-\theta)\mu^0}{\sqrt{2}n^{1.5}} \cdot \epsilon^0.$$

Hence, from Lemma 1 there is a variable indexed $\bar{j}$, such that $\bar{j} \in N^*$, and

$$s(\mu)_{\bar{j}} \geq \frac{\|\mathbf{s}^*\|_\infty}{2n} \geq \frac{\sqrt{1-\eta_0}(1-\theta)\mu^0}{2\sqrt{2}n^{2.5}} \cdot \epsilon^0,$$

for all $\mu \in (0, \mu^0]$. ∎

Now consider two cases:

$$\frac{\sqrt{n}\epsilon^0}{\eta_0} \geq 1. \tag{20}$$

and

$$\frac{\sqrt{n}\epsilon^0}{\eta_0} < 1. \tag{21}$$

In Case (20), we have $\bar{\alpha} = 0$, and

$$\epsilon^0 \geq \frac{\eta_0}{\sqrt{n}} \quad \text{and} \quad s(\mu)_{\bar{j}} \geq \frac{\eta_0\sqrt{1-\eta_0}(1-\theta)\mu^0}{2\sqrt{2}n^3},$$

where index $\bar{j} \in N^*$ is the one singled out in Lemma 8. In this case, we continue apply the predictor-corrector path-following algorithm reducing μ from μ^0. Thus, as soon as

$$\frac{\mu}{\mu^0} \leq \frac{\eta_0\sqrt{1-\eta_0}(1-\theta)}{8\sqrt{2}n^4 g},$$

we have

$$s(\mu)_{\bar{j}} \geq \frac{\eta_0\sqrt{1-\eta_0}(1-\theta)\mu^0}{2\sqrt{2}n^3} \geq 4n\mu \cdot g,$$

and at any given approximate central path point $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ such that $\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) \leq \eta_0$,

$$s_{\bar{j}} \geq \frac{4}{5}s(\mu)_{\bar{j}} \geq \frac{16n\mu \cdot g}{5} \geq \frac{8n\mu \cdot g}{3}.$$

That is, $\bar{j} \in J_3(\mu)$ and it can now be eliminated.

In Case (21), we must have

$$1 - \bar{\alpha} = \frac{\sqrt{n}\epsilon^0}{\eta_0} \quad \text{and} \quad s(\mu)_{\bar{j}} \geq \frac{\eta_0\sqrt{1-\eta_0}(1-\theta)(1-\bar{\alpha})\mu^0}{2\sqrt{2}n^3}$$

where again index $\bar{j}$ is the one singled out in Lemma 8. Note that the first predictor step has reduced μ^0 to $(1-\bar{\alpha})\mu^0$. Then, we continue apply the predictor-corrector algorithm reducing μ from $(1-\bar{\alpha})\mu^0$. As soon as

$$\frac{\mu}{(1-\bar{\alpha})\mu^0} \leq \frac{\eta_0\sqrt{1-\eta_0}(1-\theta)}{8\sqrt{2}n^4 g},$$

we have again

$$s(\mu)_{\bar{j}} \geq 4n\mu \cdot g,$$

and at any given approximate central path point $(\mathbf{x}, \mathbf{y}, \mathbf{s})$ such that $\eta(\mathbf{x}, \mathbf{y}, \mathbf{s}, \mu) \leq \eta_0$,

$$s_{\bar{j}} \geq \frac{4}{5}s(\mu)_{\bar{j}} \geq \frac{16n\mu \cdot g}{5} \geq \frac{8n\mu \cdot g}{3}.$$

That is, $\bar{j} \in J_3(\mu)$ and it can now be eliminated.

Note that for both cases the number of predictor-corrector algorithm iterations is at most $O(n^{0.5}(\log \frac{1}{1-\theta} + \log n))$ to reduce μ below the specified values such that $J_3(\mu)$ contains at least one variable in N^* and can be eliminated from the MDP. This major step can be repeated and the number of such steps should be no more than $|N^*| \leq n$ (till final $N^* = \emptyset$ and the algorithm terminates by Lemma 6). To summarize:

Theorem 1 *The combinatorial interior-point algorithm generates an optimal solution of (14) in at most n major eliminating steps, and each step uses $O(n^{0.5}(\log \frac{1}{1-\theta} + \log n))$ predictor-corrector interior-point algorithm iterations.*

Using the Karmakar rank-one updating scheme, the average number of arithmetic operations of each predictor-corrector interior-point iteration is $O(n^{2.5})$. Thus,

Theorem 2 *The combinatorial interior-point algorithm generates an optimal solution of (14) in at most $O(n^4(\log \frac{1}{1-\theta} + \log n))$ arithmetic operations.*

The similar proof can apply to the MDP with k actions for each state, where we have

Corollary 1 *The combinatorial interior-point algorithm generates an optimal solution of the MDP in at most $(k-1)n$ major eliminating steps, and each step uses $O((nk)^{0.5}(\log \frac{1}{1-\theta} + \log n + \log k))$ predictor-corrector interior-point algorithm iterations, where n is the number of states and k is the number of actions for each state. The total arithmetic operations to solve the MDP is bounded by $O((nk)^4(\log \frac{1}{1-\theta} + \log n + \log k))$.*

References

[1] D.A. Bayer and J.C. Lagarias, The nonlinear geometry of linear programming. II. Legendre transform coordinates and central trajectories, Transactions of the AMS 314 (1989) 527–581.

[2] L. Blum, F. Cucker, M. Shub, and S. Smale, Complexity and Real Computation (Springer-Verlag, 1996).

[3] G. B. Dantzig, Optimal solutions of a dynamic Leontief model with substitution, Econometrica 23 (1955) 295-302.

[4] G. B. Dantzig, A proof of a property of Leontief and Markov matrices, Operations Research Center, UC Berkeley, RR-25 (1962)

[5] G. de Ghellinck, Les Probléms de Décisions Séquentielles, Cahiers du Centre d'Etudes de Recherche Opérationnelle 2, 2 (1960) 161-179.

[6] F. D'Epenoux, A Probabilistic Production and Inventory Problem, Management Science 10, 1(1963) 98-108 (Translation of an article published in Revue Francaise de Recherche Opérationnelle 14 (1960)).

[7] R. M. Fruend and J. R. Vera, Some characterization and properties of the "distance to ill-posedness" and the condition measures of a conic linear system, Mathematical Programming 86 (1999) 225-260.

[8] C.C. Gonzaga, Path-following methods for linear programming, SIAM Review 34 (1992) 167–224.

[9] J. C. K. Ho and L. Tuncel, Reconciliation of various complexity and condition measures for linear programming problems and a generalization of Tardos' theorem, FOUNDATIONS OF COMPUTATIONAL MATHEMATICS, Proceedings of the Smalefest, World Scientific, 2002, pp. 93-147.

[10] N. Karmarkar, A new polynomial-time algorithm for linear programming, Combinatorica 4 (1984) 373–395.

[11] L.G. Khachiyan, A polynomial algorithm in linear programming, Dokl. Akad. Nauk SSSR 244 (1979) 1093–1086. Translated in *Soviet Math. Dokl.* 20:191–194.

[12] M. Kojima, S. Mizuno and A. Yoshise, A polynomial-time algorithm for a class of linear complementarity problems, Mathematical Programming 44 (1989) 1–26.

[13] M. L. Littman, T. L. Dean and L. P. Kaelbling, On the complexity of solving Markov decision problems, Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence (UAI–95), 1995, pp. 394-402.

[14] A. S. Manne, Linear programming and sequential decisions, Management Science 6 (1960) 259-267.

[15] Y. Mansour and S. Singh, On the complexity of policy iteration, Proceedings of the 15th International Conference on Uncertainty in AI, 1999, pp. 401-408.

[16] N. Megiddo, Pathways to the optimal set in linear programming, in: Progress in Mathematical Programming: Interior Point and Related Method, N. Megiddo, ed. (Springer, New York, 1989) pp. 131–158.

[17] N. Megiddo, S. Mizuno and T. Tsuchiya, A modified layered-step interior-point algorithm for linear programming, Mathematical Programming 82 (1998) 339-356.

[18] S. Mizuno, M.J. Todd and Y. Ye, On adaptive-step primal-dual interior-point algorithms for linear programming, Mathematics of Operations Research 18 (1993) 964–981.

[19] R.C. Monteiro and I. Adler, Interior path following primal-dual algorithm. Part I: linear programming, Mathematical Programming 44 (1989) 27–41.

[20] R.C. Monteiro and T. Tsuchiya, A variant of the Vavasis-Ye layered-step interior-point algorithm for linear programming, SIAM J. Optimization 13 (2003) 1054-1079.

[21] R.C. Monteiro and T. Tsuchiya, A new iteration-complexity bound for the MTY predictor-corrector algorithm, Research Memorandom No. 856, The Inistitute of Statistical Mathematics, Tokyo, Japan, October 2002 (to appear in SIAM J. Optimization).

[22] C.H. Papadimitriou and K. Steiglitz, Combinatorial Optimization: Algorithms and Complexity (Prentice Hall, Englewood Cliffs, New Jersey, 1982).

[23] J. Renegar, A polynomial-time algorithm based on Newton's method for linear programming, Mathematical Programming 40 (1988) 59–94.

[24] J. Renegar, Linear programming, complexity theory, and elementary functional analysis, Mathematical Programming 70 (1995) 279-351.

[25] B. Van Roy, Learning and value function approximation in complex decision processes, Ph.D. Thesis, MIT (1998).

[26] S. Vavasis and Y. Ye, A primal-dual interior-point method whose running time depends only on the constraint matrix, Mathematical Programming 74 (1996) 79–120.

[27] A. Veinott, Extreme points of Leontief substitution systems, Linear Algebra and its Applications 1 (1968) 181-194.

[28] Y. Ye, Interior-Point Algorithm: Theory and Analysis, Wiley-Interscience Series in Discrete Mathematics and Optimization, (John Wiley & Sons, 1997).

[29] Y. Ye, On the finite convergence of interior-point algorithms for linear programming, Mathematical Programming 57 (1992) 325-335.

[30] Y. Ye, A 'build-down' scheme for linear programming, Mathematical Programming 46 (1990) 61-72.