Joint Technical Report of U.H. and M.I.T.

Technical Report C-2008-43 *and* LIDS Report 2797
Dept. Computer Science Dept. EECS
University of Helsinki M.I.T.

July 2008

New Error Bounds for Approximations from Projected Linear Equations

Huizhen Yu*
janey.yu@cs.helsinki.fi

Dimitri P. Bertsekas†
dimitrib@mit.edu

Abstract

We consider linear fixed point equations and their approximations by projection on a low dimensional subspace. We derive new bounds on the approximation error of the solution, which are expressed in terms of low dimensional matrices and can be computed by simulation. When the fixed point mapping is a contraction, as is typically the case in Markovian decision processes (MDP), one of our bounds is always sharper than the standard worst case bounds, and another one is often sharper. Our bounds also apply to the non-contraction case, including policy evaluation in MDP with nonstandard projections that enhance exploration. There are no error bounds currently available for this case to our knowledge.

*Huizhen Yu is with HIIT and Dept. Computer Science, University of Helsinki, Finland.
†Dimitri Bertsekas is with the Laboratory for Information and Decision Systems (LIDS), M.I.T.

Contents

1 Introduction

For a given $n \times n$ matrix A and vector $b \in \Re^n$, let x^* and $\bar{x}$ be solutions of the two linear fixed point equations,

$$x = Ax + b, \qquad x = \Pi(Ax + b), \tag{1}$$

respectively, where Π denotes projection on a k-dimensional subspace S with respect to certain weighted Euclidean norm $\|\cdot\|_\xi$. We assume that x^* and $\bar{x}$ exist, and that the matrix $I - \Pi A$ is invertible so that $\bar{x}$ is unique.

Our objective in solving the projected equation $x = \Pi(Ax + b)$ is to approximate the solution of the original equation $x = Ax + b$ using k-dimensional computations and storage. Implicit here is the assumption that n is very large, so that n-dimensional vector-matrix operations are practically impossible, while $k << n$. This approach is common in approximate dynamic programming, and has been central in much of recent research on the subject (see e.g., [Sut88, TV97, BT96, SB98, Ber07]). In particular, in the context of MDP and policy iteration algorithms, the evaluation of the cost vector of a fixed policy requires solution of the equation $x = Ax+b$, where A is a stochastic or substochastic matrix. Simulation-based approximate policy evaluation methods, based on temporal differences (TD), such as TD(λ), LSTD(λ), and LSPE(λ), have been successfully used to approximate the policy cost vector by solving a projected equation $x = \Pi(Ax + b)$ with low-order computation and storage (see e.g., [Sut88, TV97, BT96, SB98, Ber07]). In our recent paper [BY08], we have extended TD-type methods to the case where A is an arbitrary matrix, subject only to the restriction that $I - \Pi A$ is invertible. In the present paper, we derive bounds on the distance/error between x^* and $\bar{x}$. Our bounds apply to the general context where A is arbitrary, but are new even when specialized to the MDP context.

In the MDP context, where ΠA is usually a contraction, there are two commonly used error bounds that compare the norms of $x^* - \bar{x}$ and $x^* - \Pi x^*$. The first bound (see e.g., [BT96, TV97]) holds if $\|\Pi A\| = \alpha < 1$ with respect to some norm $\|\cdot\|$, and has the form

$$\|x^* - \bar{x}\| \leq \frac{1}{1-\alpha}\|x^* - \Pi x^*\|. \tag{2}$$

The second bound (see e.g., [TV99a, Ber07]) holds in the usual case where ΠA is a contraction with respect to the Euclidean norm $\|\cdot\|_\xi$, with ξ being the invariant distribution of the Markov chain underlying the problem, i.e., $\|\Pi A\|_\xi = \alpha < 1$. It is derived using the Pythagorean theorem $\|x^* - \bar{x}\|_\xi^2 = \|x^* - \Pi x^*\|_\xi^2 + \|\bar{x} - \Pi x^*\|_\xi^2$, and it is much sharper than the first bound:

$$\|x^* - \bar{x}\|_\xi \leq \frac{1}{\sqrt{1-\alpha^2}}\|x^* - \Pi x^*\|_\xi. \tag{3}$$

The bounds (2), (3) are determined by the modulus of contraction α, and apply only when ΠA is a contraction mapping. We develop in this paper new error bounds, which are sharper when ΠA is a contraction, including important MDP cases, and also apply when ΠA is not a contraction.

Our starting point is the observation that the two terms involved in the bounds (2) and (3) satisfy the following equation with or without contraction assumptions:[1]

$$x^* - \bar{x} = (I - \Pi A)^{-1}(x^* - \Pi x^*). \tag{4}$$

We may view the bounds (2), (3) as relaxed versions of this equation. In particular, we may obtain the bound (2) by writing

$$(I - \Pi A)^{-1} = I + \Pi A + \cdots,$$

[1] This can be seen by subtracting $\bar{x} = \Pi(A\bar{x} + b)$ from $\Pi x^* = \Pi(Ax^* + b)$ to obtain

$$\Pi x^* - \bar{x} = \Pi A(x^* - \bar{x}), \quad \Rightarrow \quad (\Pi x^* - x^*) + (x^* - \bar{x}) = \Pi A(x^* - \bar{x}), \quad \Rightarrow \quad (4).$$

and by upper-bounding each term in the expansion separately: $\|(\Pi A)^n\| \leq \alpha^n$. We may obtain the bound (3) by writing

$$(I - \Pi A)^{-1} = I + \Pi A(I - \Pi A)^{-1}, \tag{5}$$

and by upper-bounding the norm of $\Pi A(I - \Pi A)^{-1}(x^* - \Pi x^*)$ by $\alpha\|x^* - \bar{x}\|_\xi$ and rearranging terms.[2] We will develop a different bounding approach, so that α will not be in the denominator of the bound. To this end, we will express $(I - \Pi A)^{-1}$ in the form

$$(I - \Pi A)^{-1} = I + (I - \Pi A)^{-1}\Pi A, \tag{6}$$

and aim at bounding the term $(I - \Pi A)^{-1}\Pi A(x^* - \Pi x^*)$ *directly* (this term is in fact $\Pi x^* - \bar{x}$, the bias of $\bar{x}$ from Πx^*). In doing so, we will obtain bounds that not only can be sharper than the preceding bounds for the contraction case, but also carry over to the non-contraction case.

We will derive two bounds, which involve the spectral radii of small-size matrices, and provide a "data/problem-dependent" error analysis, in contrast to the fixed error bounds (2), (3); see Theorems 1 and 2. The bounds are *independent* of the parametrization of the subspace S, and can be computed with low-dimensional operations and simulation, if this is desirable. One of the bounds is sharper than the other, but involves more complex computations. We also derive some additional bounds that provide insight into the character of the approximation error, but are qualitative in nature; see Props. 3 and 4.

Most of our bounds have the general form

$$\|x^* - \bar{x}\|_\xi \leq B(A, \xi, S)\, \|x^* - \Pi x^*\|_\xi, \tag{7}$$

where $B(A, \xi, S)$ is a constant that depends on A, ξ, and S (but not on b). Like the bounds (2), (3), we may view $\|x^* - \Pi x^*\|_\xi$ as the *baseline error*, i.e., the minimum error in estimating x^* by a vector in the approximation subspace S. We may view $B(A, \xi, S)$ as an upper bound to the *amplification ratio*

$$\frac{\|x^* - \bar{x}\|_\xi}{\|x^* - \Pi x^*\|_\xi},$$

which is due to solving the projected equation $x = \Pi(Ax + b)$ instead of projecting x^* on S, or equivalently, view $\sqrt{B^2(A, \xi, S) - 1}$ as an upper bound to the *"bias-to-distance" ratio*

$$\frac{\|\bar{x} - \Pi x^*\|_\xi}{\|x^* - \Pi x^*\|_\xi}.$$

Figure 1 illustrates this relation between the bound, x^* and $\bar{x}$.

We present our main results in the next section. In Section 3, we address the application of the new error bounds to the approximate policy evaluation in MDP and to the far more general problem of approximate solution of large systems of linear equations. In Section 4, we present additional related results based on the same line of analysis, including improved qualitative bounds, as well as analogous computable error bounds for a different approximation method: the equation error minimization approach.

2 Main Results

We first introduce the main theorems and explain the underlying ideas, and then give the proofs in Section 2.1. Let Φ be an $n \times k$ matrix whose columns form a basis of S. Let Ξ be a diagonal matrix with the components of ξ on the diagonal. Define $k \times k$ matrices B, M, and F by

$$B = \Phi'\Xi\Phi, \qquad M = \Phi'\Xi A\Phi, \qquad F = (I - B^{-1}M)^{-1} \tag{8}$$

[2] From Eqs. (4)-(5) and the orthogonality of $(x^* - \Pi x^*)$ to the subspace S, we have

$$\begin{aligned}\|x^* - \bar{x}\|_\xi^2 &= \|x^* - \Pi x^*\|_\xi^2 + \|\Pi A(I - \Pi A)^{-1}(x^* - \Pi x^*)\|_\xi^2 \\ &= \|x^* - \Pi x^*\|_\xi^2 + \|\Pi A(x^* - \bar{x})\|_\xi^2 \leq \|x^* - \Pi x^*\|_\xi^2 + \alpha^2\|x^* - \bar{x}\|_\xi^2.\end{aligned}$$

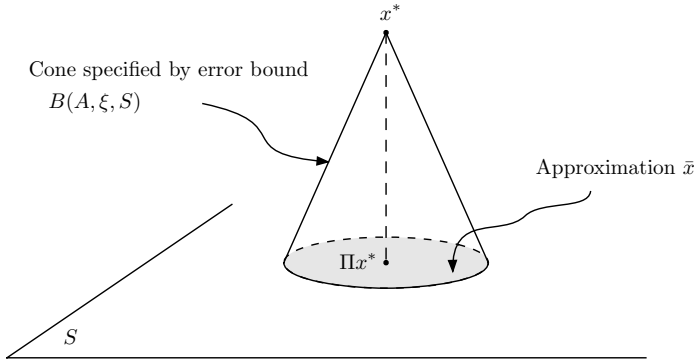


Figure 1: The relation between the error bound and $\bar{x}$: $\bar{x}$ lies in the intersection of S and a cone which originates from x^* and whose angle is specified by the error bound $B(A,\xi,S)$ as $\cos^{-1}\left(\frac{1}{B(A,\xi,S)}\right)$. The smaller $B(A,\xi,S)$ is, the shaper the cone. The smallest bound $B(A,\xi,S)=1$ implies $\bar{x}=\Pi x^*$.

(we will show later that the inverse in the definition of F exists). Notice that the projection matrix Π can be expressed as $\Pi=\Phi(\Phi'\Xi\Phi)^{-1}\Phi'\Xi=\Phi B^{-1}\Phi'\Xi$. For a square matrix L, let $\sigma(L)$ denote the spectral radius of L.

Throughout the paper, x^* denotes some solution of the equation $x=Ax+b$; we implicitly assume that such a solution exists. When reference is made to $\bar{x}$, we implicitly assume that $I-\Pi A$ is invertible, and that $\bar{x}$ is the unique solution of the equation $x=\Pi(Ax+b)$.

Theorem 1. *The approximation error $x^*-\bar{x}$ satisfies*

$$\|x^*-\bar{x}\|_\xi \le \sqrt{1+\sigma(G_1)\|A\|_\xi^2}\,\|x^*-\Pi x^*\|_\xi, \tag{9}$$

where G_1 is the $k\times k$ matrix

$$G_1=B^{-1}F'BF. \tag{10}$$

Furthermore,

$$\sigma(G_1)=\|(I-\Pi A)^{-1}\Pi\|_\xi^2,$$

so the bound (9) is invariant to the choice of basis vectors of S (i.e., Φ).

The idea in deriving Theorem 1 is to combine Eqs. (4)-(5) with the bound

$$\left\|(I-\Pi A)^{-1}\Pi A(x^*-\Pi x^*)\right\|_\xi \le \left\|(I-\Pi A)^{-1}\Pi\right\|_\xi \|A\|_\xi \|x^*-\Pi x^*\|_\xi,$$

and to show that $\|(I-\Pi A)^{-1}\Pi\|_\xi^2=\sigma(G_1)$. An important fact, to be demonstrated later, is that G_1 can be obtained by simulation, using low dimensional calculations.

While the bound of Theorem 1 can be conveniently computed, it is less sharp than the bound of the subsequent Theorem 2, and under certain circumstances less sharp than the bound (3). In Theorem 1, $\|A\|_\xi$ is needed, and this can be a drawback, particularly for the non-contraction case. In Theorem 2, $\|A\|_\xi$ is no longer needed; A is absorbed into the matrix to be estimated. Furthermore, Theorem 2 takes into account that $x^*-\Pi x^*$ is perpendicular to the subspace S; this considerably sharpens the bound. On the other hand, the sharpened bound of Theorem 2 involves a $k\times k$ matrix R (defined below) in addition to B and M, which may not be straightforward to estimate in some cases, as will be commented later.

Theorem 2. *The approximation error $x^*-\bar{x}$ satisfies*

$$\|x^*-\bar{x}\|_\xi \le \sqrt{1+\sigma(G_2)}\,\|x^*-\Pi x^*\|_\xi, \tag{11}$$

where G_2 is the $k\times k$ matrix

$$G_2=B^{-1}F'BFB^{-1}(R-MB^{-1}M'), \tag{12}$$

and R is the $k \times k$ matrix

$$R = \Phi' \Xi A \Xi^{-1} A' \Xi \Phi.$$

Furthermore,

$$\sigma(G_2) = \|(I - \Pi A)^{-1} \Pi A (I - \Pi)\|_\xi^2,$$

so the bound (11) is invariant to the choice of basis vectors of S (i.e., Φ).

The idea in deriving Theorem 2 is to combine Eqs. (4)-(5) with the bound

$$\begin{aligned} \left\|(I - \Pi A)^{-1} \Pi A (x^* - \Pi x^*)\right\|_\xi &= \left\|(I - \Pi A)^{-1} \Pi A (I - \Pi)(x^* - \Pi x^*)\right\|_\xi \\ &\leq \left\|(I - \Pi A)^{-1} \Pi A (I - \Pi)\right\|_\xi \left\|x^* - \Pi x^*\right\|_\xi, \end{aligned}$$

and to show that $\|(I-\Pi A)^{-1}\Pi A(I-\Pi)\|_\xi^2 = \sigma(G_2)$. Incorporating the matrix $I-\Pi$ in the definition of G_2 is crucial for improving the bound of Theorem 1.

Estimating the matrix R, although not always as straightforward as estimating B and M, can be done for a number of applications. A primary exception is when A itself is an infinite sum of powers of matrices, which is the case of the TD(λ) method with $\lambda > 0$. We will address these issues in Section 2.3.

2.1 Proofs of Theorems

We shall need two technical lemmas. The first lemma introduces an expression of the matrix $(I - \Pi A)^{-1}$ that will be used to derive our error bounds. The second lemma establishes the relation between the norm of an $n \times n$ matrix that is a product of $n \times k$ and $k \times n$ matrices, and the spectral radius of a certain product of $k \times k$ matrices.

Lemma 1. *The matrix $I - \Pi A$ is invertible if and only if the inverse $(I - B^{-1}M)^{-1}$ defining F exists. When $I - \Pi A$ is invertible, $(I - \Pi A)^{-1}$ maps S onto S, and furthermore,*

$$(I - \Pi A)^{-1} = I + (I - \Pi A)^{-1} \Pi A = I + \Phi F B^{-1} \Phi' \Xi A. \tag{13}$$

Proof. We prove the second part first. For any $y \in S$, $(I - \Pi A)^{-1} y$ is the unique solution of the equation $x = \Pi A x + y$, so it lies in S. Since $(I - \Pi A)^{-1}$ has full rank, this shows that $(I - \Pi A)^{-1}$ maps S onto S.

Since $(I - \Pi A)^{-1}$ maps S onto S, we have

$$(I - \Pi A)^{-1} \Phi = \Pi (I - \Pi A)^{-1} \Phi. \tag{14}$$

Furthermore, since Φ (whose columns form a basis of S) defines a one-to-one correspondence between $\Re^k$ and S, with the inverse mapping given by $B^{-1}\Phi'\Xi$ (as can be seen from the expression of Π), the following three-mapping composition,

$$H = (B^{-1}\Phi'\Xi) \cdot (I - \Pi A)^{-1} \cdot \Phi,$$

is a one-to-one mapping from $\Re^k \to \Re^k$. It follows that two vectors $v, r \in \Re^k$ satisfy $Hv = r$ if and only if $(I - \Pi A)^{-1}\Phi v = \Phi r$, or equivalently if and only if $\Phi r = \Pi A \Phi r + \Phi v$, or equivalently, if and only if $r = B^{-1}\Phi'\Xi A \Phi r + v$. Using the definitions of M and F, this implies that

$$H = (I - B^{-1}\Phi'\Xi A\Phi)^{-1} = (I - B^{-1}M)^{-1} = F. \tag{15}$$

From Eqs. (14) and (15), and the expression of Π, we have

$$\begin{aligned} (I - \Pi A)^{-1}\Pi &= \Pi (I - \Pi A)^{-1} \Pi \\ &= \Phi (B^{-1}\Phi'\Xi)(I - \Pi A)^{-1} \Phi B^{-1} \Phi' \Xi \\ &= \Phi H B^{-1} \Phi' \Xi \\ &= \Phi F B^{-1} \Phi' \Xi, \end{aligned} \tag{16}$$

and right-multiplying both sides by A and adding I, we obtain Eq. (13).

We now prove the first part. If $I - \Pi A$ is invertible, the proof preceding Eq. (15) shows that $(I - B^{-1}M)^{-1}$ exists. Conversely, if $(I - B^{-1}M)^{-1}$ exists, the argument immediately preceding Eq. (15) shows that $I - \Pi A$ is a one-to-one mapping on S and therefore cannot have $z \neq 0$ such that $\Pi A z = z$. This shows that 1 is not an eigenvalue of ΠA, so $I - \Pi A$ is invertible. □

Remark 1. Note that since B and M are low-dimensional matrices, the first part of Lemma 1 is useful for verifying the existence of the inverse of $I - \Pi A$ using the data.

Lemma 2. *Let H and D be an $n \times k$ and $k \times n$ matrix, respectively. Let $\|\cdot\|$ denote the standard (unweighted) Euclidean norm. Then,*

$$\|HD\|_\xi^2 = \|\Xi^{1/2} H D \Xi^{-1/2}\|^2 = \sigma\big((H'\Xi H)(D\Xi^{-1}D')\big). \tag{17}$$

Proof. By the definition of $\|\cdot\|_\xi$, for any $x \in \Re^n$, $\|x\|_\xi = \|\Xi^{1/2}x\|$, where $\|\cdot\|$ is the standard Euclidean norm. The first equality in Eq. (17) then follows from the definition of the norms: for any $n \times n$ matrix E,

$$\begin{aligned}\|E\|_\xi &= \sup_{\|x\|_\xi = 1} \|Ex\|_\xi = \sup_{\|\Xi^{1/2}x\| = 1} \|\Xi^{1/2} E x\| \\ &= \sup_{\|z\|=1} \|\Xi^{1/2} E \Xi^{-1/2} z\| = \|\Xi^{1/2} E \Xi^{-1/2}\|,\end{aligned}$$

where a change of variable $z = \Xi^{1/2}x$ is applied to derive the first equality in the second line.

For a square matrix E, we have $\|E\| = \sqrt{\sigma(E'E)}$. Letting $E = \Xi^{1/2} H D \Xi^{-1/2}$, we proceed to prove the second equality in Eq. (17), by studying the spectral radius of the symmetric positive semidefinite matrix $E'E$. Define $W = H'\Xi H$ to simplify notation. We have,

$$E'E = \Xi^{-1/2} D' H' \Xi^{1/2} \cdot \Xi^{1/2} H D \Xi^{-1/2} = \Xi^{-1/2} D' W D \Xi^{-1/2}.$$

Let λ be a nonzero (necessarily real) eigenvalue of $E'E$, and let x be a nonzero corresponding eigenvector. We have

$$\Xi^{-1/2} D' W D \Xi^{-1/2} x = \lambda x, \tag{18}$$

so x is in $\mathrm{col}(\Xi^{-1/2}D')$ and can be expressed as

$$x = \Xi^{-1/2} D' \bar{r}$$

for some vector $\bar{r} \in \Re^k$. Let

$$r = \frac{1}{\lambda} W D \Xi^{-1/2} x = \frac{1}{\lambda} W D \Xi^{-1} D' \bar{r}.$$

Then, by Eq. (18),

$$\Xi^{-1/2} D' r = \frac{\lambda}{\lambda} x = \Xi^{-1/2} D' \bar{r}, \quad \Rightarrow \quad D' r = D' \bar{r},$$

thus,

$$\lambda r = W D \Xi^{-1} D' \bar{r} = W D \Xi^{-1} D' r. \tag{19}$$

This implies that λ and r are an eigenvalue-eigenvector pair of the matrix $W(D\Xi^{-1}D')$. Conversely, it is easy to see that if λ and r are an eigenvalue-eigenvector pair of the matrix $W(D\Xi^{-1}D')$, then λ and $\Xi^{-1/2}D'r$ are an eigenvalue-eigenvector pair of the matrix $E'E$. Therefore,

$$\sigma\big(E'E\big) = \sigma\big(W(D\Xi^{-1}D')\big) = \sigma\big((H'\Xi H)(D\Xi^{-1}D')\big),$$

proving the second equality in Eq. (17). □

We now proceed to prove Theorem 1.

Proof of Theorem 1. To simplify notation, let us denote $y = x^* - \Pi x^*$ and $C = FB^{-1}$. By Lemma 1,

$$(I - \Pi A)^{-1} y = y + \Phi C \Phi' \Xi A y,$$

and since y is orthogonal to S and the second term on the right-hand-side lies in S, by the Pythagorean theorem, we have

$$\|(I - \Pi A)^{-1} y\|_\xi^2 = \|y\|_\xi^2 + \|\Phi C \Phi' \Xi A y\|_\xi^2. \tag{20}$$

Applying Lemma 2 to the matrix $\Phi C \Phi' \Xi$ with $H = \Phi C$ and $D = \Phi' \Xi$ and denoting by G the matrix $(H' \Xi H)(D \Xi^{-1} D')$, the second term on the right-hand-side of Eq. (20) can be bounded by

$$\begin{aligned} \|\Phi C \Phi' \Xi A y\|_\xi &\leq \|\Phi C \Phi' \Xi\|_\xi \, \|A y\|_\xi \\ &= \sqrt{\sigma(G)} \, \|A y\|_\xi \\ &\leq \sqrt{\sigma(G)} \, \|A\|_\xi \, \|y\|_\xi. \end{aligned} \tag{21}$$

We have

$$G = (C' \Phi' \Xi \Phi C)(\Phi' \Xi \Xi^{-1} \Xi \Phi) = (FB^{-1})' B (FB^{-1}) B = B^{-1} F' B F,$$

so G is the matrix G_1 given in the statement of the theorem.

By combining Eq. (4), and Eqs. (20) and (21), it follows that

$$\|x^* - \bar{x}\|_\xi^2 \, \leq \, \big(1 + \sigma(G_1) \|A\|_\xi^2\big) \, \|x^* - \Pi x^*\|_\xi^2,$$

which proves the bound (9).

Finally, tracing the proof argument backwards, we see that $\sigma(G_1) = \|\Phi F B^{-1} \Phi' \Xi\|_\xi^2$, while by Eq. (16) given in the proof of Lemma 1,

$$\Phi F B^{-1} \Phi' \Xi = (I - \Pi A)^{-1} \Pi.$$

Thus, $\sigma(G_1)$ is equal to $\|(I - \Pi A)^{-1} \Pi\|_\xi^2$, and depends only on S and ξ and not the choice of Φ. This completes the proof. □

We now prove Theorem 2.

Proof of Theorem 2. Let us denote $y = x^* - \Pi x^*$ and $C = FB^{-1}$. As shown in the proof of Theorem 1,

$$\|(I - \Pi A)^{-1} y\|_\xi^2 = \|y\|_\xi^2 + \|\Phi C \Phi' \Xi A y\|_\xi^2. \tag{22}$$

We proceed to bound the second term. Since

$$(I - \Pi)(x^* - \Pi x^*) = x^* - \Pi x^*,$$

i.e., $(I - \Pi) y = y$, we have

$$\|\Phi C \Phi' \Xi A y\|_\xi = \|\Phi C \Phi' \Xi A (I - \Pi) y\|_\xi \leq \|\Phi C \Phi' \Xi A (I - \Pi)\|_\xi \|y\|_\xi. \tag{23}$$

Applying Lemma 2 to the matrix $\Phi C \Phi' \Xi A (I - \Pi)$ with $H = \Phi C$ and $D = \Phi' \Xi A (I - \Pi)$, and denoting by G the matrix $(H' \Xi H)(D \Xi^{-1} D')$, we have

$$\|\Phi C \Phi' \Xi A (I - \Pi)\|_\xi = \sqrt{\sigma(G)}. \tag{24}$$

We now verify that the matrix $G = (H' \Xi H)(D \Xi^{-1} D')$ is the matrix G_2 given in the statement of the theorem. It can be seen that

$$H' \Xi H = C' B C, \qquad D \Xi^{-1} D' = \Phi' \Xi A (I - \Pi) \Xi^{-1} (I - \Pi)' A' \Xi \Phi.$$

Since $\Pi\Xi^{-1} = \Phi B^{-1}\Phi'\Xi\Xi^{-1} = \Phi B^{-1}\Phi'$, we have

$$\begin{aligned}(I-\Pi)\Xi^{-1}(I-\Pi)' &= \Xi^{-1} - \Pi\Xi^{-1} - \Xi^{-1}\Pi' + \Pi\Xi^{-1}\Pi' \\ &= \Xi^{-1} - 2\Phi B^{-1}\Phi' + \Phi B^{-1}\Phi'\Xi\Phi B^{-1}\Phi' \\ &= \Xi^{-1} - \Phi B^{-1}\Phi'.\end{aligned}$$

So the matrix $D\Xi^{-1}D'$ is

$$\begin{aligned}\Phi'\Xi A(I-\Pi)\Xi^{-1}(I-\Pi)'A'\Xi\Phi &= \Phi'\Xi A\big(\Xi^{-1} - \Phi B^{-1}\Phi'\big)A'\Xi\Phi \\ &= \Phi'\Xi A\Xi^{-1}A'\Xi\Phi - \Phi'\Xi A\Phi B^{-1}\Phi'A'\Xi\Phi \\ &= R - MB^{-1}M'\end{aligned}$$

with $R = \Phi'\Xi A\Xi^{-1}A'\Xi\Phi$, and the matrix

$$G = C'BC(D\Xi^{-1}D') = (FB^{-1})'B(FB^{-1})(R - MB^{-1}M')$$

is the matrix G_2 given in the statement.

The rest of the proof is similar to that of Theorem 1: we use Eqs. (4) and (22)-(24) to establish the bound, and we trace the proof argument backwards to establish that $\sqrt{\sigma(G_2)} = \|(I-\Pi A)^{-1}\Pi A(I-\Pi)\|_\xi$. □

Remark 2. The same line of analysis applies in the case where the weights defining the Euclidean projection Π are different from ξ, the weights defining the norm which is used to evaluate the approximation quality. In such a case, we use the triangle inequality in place of the Pythagorean theorem; the bounds are similarly expressed in terms of small size matrices, and with additional care, they can also be estimated by simulation.

2.2 Comparison of Error Bounds

The error bounds of Theorems 1 and 2 apply to the general case where ΠA is not necessarily a contraction mapping, while the worst case error bounds (2) and (3) only apply when ΠA is a contraction. We will thus compare them for the contraction case. Nevertheless, our discussion will illuminate the strengths and weaknesses of the new bounds for both contraction and non-contraction cases.

First we show that the error bound of Theorem 2 is always the sharpest.

Proposition 1. *Assume that $\|\Pi A\|_\xi \le \alpha < 1$. Then, the error bound of Theorem 2 is always no worse than the error bound (3), i.e.,*

$$1 + \sigma(G_2) \le \frac{1}{1-\alpha^2},$$

where G_2 is given by Eq. (12).

Proof. Let $\gamma = \sqrt{\sigma(G_2)}$. Since $\sigma(G_2) = \|(I-\Pi A)^{-1}\Pi A(I-\Pi)\|_\xi^2$ by Theorem 2, what we need to show is that

$$\gamma^2 = \|(I-\Pi A)^{-1}\Pi A(I-\Pi)\|_\xi^2 \le \frac{1}{1-\alpha^2} - 1 = \frac{\alpha^2}{1-\alpha^2}.$$

Consider a vector $y \neq 0$ such that

$$\|(I-\Pi A)^{-1}\Pi A(I-\Pi)y\|_\xi = \gamma\|y\|_\xi. \tag{25}$$

Since γ equals the matrix norm, we must have $(I-\Pi)y = y$, i.e., $\Pi y = 0$. (Otherwise, by redefining y to be $y - \Pi y$, we can decrease $\|y\|_\xi$ while keeping the value of the left hand side of (25) unchanged, which would imply an increase in γ, a contradiction.) Consider the two equations of x,

$$x = (y - Ay) + Ax, \qquad x = \Pi(y - Ay) + \Pi Ax = \Pi Ax - \Pi Ay.$$

Then, y is a solution of the first equation. Denote the solution of the second projected equation by $\bar{x}$. The error bound (3) implies that

$$\|\Pi y - \bar{x}\|_\xi^2 \leq \left(\frac{1}{1-\alpha^2} - 1\right) \|y - \Pi y\|_\xi^2 = \frac{\alpha^2}{1-\alpha^2}\|y - \Pi y\|_\xi^2, \tag{26}$$

while by the definition of $\bar{x}$ and y, we have

$$\Pi y - \bar{x} = -\bar{x} = (I - \Pi A)^{-1}\Pi A y = (I - \Pi A)^{-1}\Pi A(I - \Pi)y, \tag{27}$$

and by Eq. (25),

$$\|\Pi y - \bar{x}\|_\xi = \gamma\|y\|_\xi = \gamma\|y - \Pi y\|_\xi.$$

Together with Eq. (26), this implies $\gamma^2 \leq \frac{\alpha^2}{1-\alpha^2}$. □

Remark 3. The proof shows that for both contraction and non-contraction cases, the bound of Theorem 2 is tight, in the sense that for any A and S, there exists a worst case choice of b for which the bound holds with equality. This can be seen from the construction of an equation and its projected form immediately following Eq. (25).

Let us compare now the error bound of Theorem 1 with the bounds (2) and (3) from the worst case viewpoint. Since Theorem 1 is effectively equivalent to

$$\left\|(I - \Pi A)^{-1}\Pi A(x^* - \Pi x^*)\right\|_\xi \leq \left\|(I - \Pi A)^{-1}\Pi\right\|_\xi \|A\|_\xi \left\|x^* - \Pi x^*\right\|_\xi,$$

we see that the bound of Theorem 1 is never worse than the bound (2), because we have bounded the norm of the matrix $(I - \Pi A)^{-1}\Pi$ as a whole, instead of bounding each term in its expansion separately as in the case in the bound (2). However, the bound of Theorem 1 can be degraded by two over-relaxations:

(i) The residual vector $x^* - \Pi x^*$ is special, in that it satisfies $\Pi(x^* - \Pi x^*) = 0$, but the bound does not use this fact.

(ii) When ΠA is zero or near zero, the bound cannot fully utilize this fact.

The effect of (i) can be quite significant when A has a dominant real eigenvalue β with an eigenvector x that lies in the approximation subspace S. In such a case, the bound reduces essentially to the bound (2), since

$$\|(I - \Pi A)^{-1}\Pi x\|_\xi = \frac{1}{1-\beta}\|x\|_\xi. \tag{28}$$

This happens because the analysis has not taken into account that the residual vector $(x^* - \Pi x^*)$ cannot be an eigenvector that is contained in S.

The relaxation related to (ii) may not look obvious in the current analysis; it does, however, in an alternative equivalent form of the analysis, by noticing that

$$(I - \Pi A)^{-1}\Pi A = \Pi A + \Pi A(I - \Pi A)^{-1}\Pi A, \tag{29}$$

and the norm of the matrix on the right has been bounded by $\|\Pi + \Pi A(I - \Pi A)^{-1}\Pi\|_\xi \|A\|_\xi$ in Theorem 1. When $\Pi A = 0$ the matrix of Eq. (29) is zero but its bound is not, because the matrices Π and A are split in the bounding procedure. Accordingly, the spectral radius $\sigma(G_1)$ becomes $\|\Pi\|_\xi^2 = 1$. Similarly, over-relaxation occurs when ΠA is not zero but is near zero.[3]

The two shortcomings of the bound of Theorem 1 arise in the MDP applications that we will discuss, as well as in non-contraction cases. On the other hand, there are cases where Theorem 1 provides sharper bounds than the fixed error bound (3), and cases where Theorem 1 gives computable

[3] In practice, when using the bound of Theorem 1, one may check if ΠA is near zero by checking if M is.

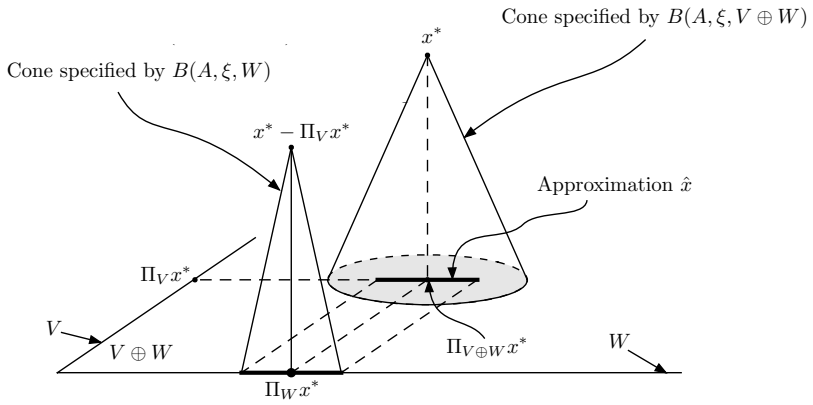


Figure 2: Illustration of Prop. 2 on transferring error bounds on one approximation subspace to another. The subspaces V and W are such that $V \perp W$ and $\Pi_V x^*$ is known. Error bounds of Theorems 1 and 2 associated with the approximation subspace W can be transfered to $V \oplus W$ by solving the projected form of an equation satisfied by $x^* - \Pi_V x^*$ with the approximation subspace being W, adding to this solution $\Pi_V x^*$, and then taking the combined solution as the approximation $\hat{x}$. In particular, $\hat{x} = \Pi_V x^* + \bar{x}_w$, where $\bar{x}_w$ is the solution of $x = \Pi_W A x + \Pi_W \tilde{b}$ with $\tilde{b} = b + A\Pi_V x^* - \Pi_V x^*$.

bounds while the bound (3) is qualitative (for example, when the modulus of contraction of ΠA is unknown). In Section 4, we will use the same line of analysis to derive strengthened versions of Theorem 1, which in part address the shortcomings just discussed.

The advantage that the bound of Theorem 1 holds over the one of Theorem 2 is that it is rather easy to compute: the matrices B and M define the solution $\bar{x}$, so the bound is obtained together with the approximating solution without extra computation overhead. By contrast, the bound of Theorem 2 involves the matrix R, which can be hard to estimate for certain applications.

We now address another way of applying Theorems 1 and 2. It is motivated by the preceding discussion on the over-relaxation (i) in the bound of Theorem 1, and it will be particularly useful for obtaining sharper bounds from Theorem 1 when the approximation subspace nearly contains eigenvectors of A associated with eigenvalues that are close to 1. The idea is to approximate the projection of x^* on a smaller subspace excluding the troublesome eigenspace and to transfer the corresponding error bound, hopefully a better bound, to the original subspace. We give a formal statement in the following proposition; see Figure 2 for an illustration. For a subspace V, let Π_V denote the projection on V.

Proposition 2. *Let V and W be two orthogonal subspaces. Assume that $\Pi_V x^*$ is known and $I - \Pi_W A$ is invertible. Let $B(A, \xi, W)$ correspond to either the error bound of Theorem 1 or that of Theorem 2 with $S = W$. Then*

$$\|x^* - \hat{x}\|_\xi \leq B(A, \xi, W)\|x^* - \Pi_{V \oplus W} x^*\|_\xi,$$

where $\hat{x} = \Pi_V x^* + \bar{x}_w$ and $\bar{x}_w$ is the solution of

$$x = \Pi_W A x + \Pi_W \tilde{b}$$

with $\tilde{b} = b + A\Pi_V x^* - \Pi_V x^*$.

Proof. First, notice that the error bounds of Theorems 1 and 2 do not depend on b. Since $x^* - \Pi_V x^*$ satisfies the linear equation $x = Ax + \tilde{b}$ with $\tilde{b} = b + A\Pi_V x^* - \Pi_V x^*$, and $\bar{x}_w$ is the solution of the corresponding projected equation, we have

$$\|(x^* - \Pi_V x^*) - \bar{x}_w\|_\xi \leq B(A, \xi, W)\|(x^* - \Pi_V x^*) - \Pi_W (x^* - \Pi_V x^*)\|_\xi.$$

Since $W \perp V$, $\Pi_W x^* = \Pi_W(x^* - \Pi_V x^*)$ and $\Pi_{V\oplus W} x^* = \Pi_V x^* + \Pi_W x^*$, therefore the above inequality is equivalent to

$$\|x^* - \hat{x}\|_\xi \le B(A, \xi, W)\|x^* - \Pi_{V\oplus W} x^*\|_\xi$$

with $\hat{x} = \Pi_V x^* + \bar{x}_w$. □

Remark 4. When V is an eigenspace of A, $A\Pi_V x^* \in V$, so $\Pi_W \tilde{b} = \Pi_W b$ by the mutual orthogonality of V and W, and $\Pi_V x^*$ is not needed in the projected equation for $\bar{x}_w$. Then, we may not need to compute $\Pi_V x^*$. An example is policy evaluation in MDP where V is the span of the constant vector of all ones. Then, $\Pi_V x^*$ is constant over all states and can be neglected in the process of policy iteration.

Remark 5. Prop. 2 also holds with $\Pi_V x^*$ replaced by any vector $v \in V$. In particular, we have

$$\|x^* - \hat{x}\|_\xi \le B(A, \xi, W)\|x^* - (v + \Pi_W x^*)\|_\xi,$$

where $\hat{x} = v + \bar{x}_w$ and $\bar{x}_w$ is the solution of the projected equation $x = \Pi_W A x + \Pi_W \tilde{b}$ with $\tilde{b} = b + Av - v$. This implication can be useful when $\Pi_V x^*$ is unknown: we may substitute v as a guess of $\Pi_V x^*$.

2.3 Estimating the Low Dimensional Matrices in the Bounds

We consider estimating the $k \times k$ matrices involved in the bounds by simulation, and we focus on estimating the matrix R in Theorem 2:

$$R = \Phi' \Xi A \Xi^{-1} A' \Xi \Phi.$$

Other cases do not seem to need explanations: the estimation of B and M using simulation has been well explained in the literature (see e.g., [Boy99, NB03, BY08]); and if instead of using simulation, products of $k \times n$ and $n \times n$ matrices can be computed directly, then the calculation of R may be done directly with common matrix algebra.

First, let us note that when the matrix Φ actually used in the simulation does not have full rank, Theorems 1 and 2 imply that the bounds can be computed by using the pseudo-inverse of B, neglecting zero eigenvalues (a tolerance level/threshold needs to be determined, of course, in the simulation context).

Without loss of generality, in this subsection, we assume that $\sum_{i=1}^n \xi_i = 1$ so that ξ can be viewed as a distribution. In practice, we never need to normalize ξ as the normalization constant will be canceled in the product defining the matrices G_1 and G_2. Let $\phi(i)'$ denote the i-th row of Φ. Our methods for estimating R are based on a common procedure: we first express R as a summation of $k \times k$ matrices, e.g.,

$$R = \sum_{i,j,\hat{j}} (a_{ji} a_{\hat{j}i}) \cdot \frac{\xi_j \xi_{\hat{j}}}{\xi_i} \cdot \phi(j)\phi(\hat{j})',$$

and guided by this expression, we generate samples and choose proper weights for them, so that each term in the summation is matched by a weighted long-run average of respective samples.

We will give four examples that apply to different contexts, depending on whether the entries of ξ and A in the preceding formula for R are explicitly known or not, with two main applications in our mind:

(i) *General linear equations* in which we know explicitly the entries of A, and we may want to choose a particular projection norm, for instance, the standard Euclidean norm (all entries of ξ being equal). The procedure of Example 1 and its slight variant in Example 2 refer primarily to this case.

(ii) *Markov decision processes* in which we do not know A, but we can generate samples by simulation of a certain Markov chain underlying the problem. Examples 3 and 4 are mostly relevant to this case, including in particular, evaluating the cost or Q-factors of a policy using TD(0)-like algorithms, with and without exploration enhancements. (We refer to our paper [BY08] for some algorithms involving exploration, where the simulation procedures of Examples 3 and 4 may apply.)

Example 1. Both ξ and A are known explicitly. We express R as the summation given above and generate a sequence of triple indices $(i_t, j_t, \hat{j}_t)$ as follows. We generate the sequence $(i_0, i_1, \ldots)$ so that its empirical distribution converges to ξ. At i_t, we generate two mutually independent transitions (i_t, j_t) and $(i_t, \hat{j}_t)$ according to a certain transition probability matrix P with $p_{ij} \neq 0$ whenever $a_{ji} \neq 0$. We then define R_t by

$$R_t = \frac{1}{t+1} \sum_{m=0}^{t} \left(\frac{a_{j_m i_m}}{p_{i_m j_m}} \cdot \frac{a_{\hat{j}_m i_m}}{p_{i_m \hat{j}_m}} \right) \cdot \frac{\xi_{j_m} \xi_{\hat{j}_m}}{\xi_{i_m}^2} \cdot \phi(j_m)\phi(\hat{j}_m)',$$

where t is a suitably large number, and approximate R by the symmetrized matrix $(R_t + R_t')/2$. Note that in the special case where $\Xi = \frac{1}{n} I$, the indices i_t can be generated independently with the uniform distribution, R reduces to $\frac{1}{n}\Phi' A A' \Phi$, and the ratio $\frac{\xi_{j_m} \xi_{\hat{j}_m}}{\xi_{i_m}^2}$ in R_t reduces to 1. □

Example 2. The weight vector ξ is not known explicitly, but A is; moreover, a sequence $(i_0, i_1, \ldots)$ can be generated so that its empirical distribution converges to ξ. For example, ξ may be the unique invariant distribution of a Markov chain, which is used to generate the sequence $(i_0, i_1, \ldots)$. In this case, we can keep tracking the empirical distribution $\hat{\xi}_t$ of the sequence i_t up to time t. We then apply the same sampling and estimation schemes as in Example 1, replacing the ratio $\frac{\xi_{j_m} \xi_{\hat{j}_m}}{\xi_{i_m}^2}$ in R_t by $\frac{\hat{\xi}_{t,j_m} \hat{\xi}_{t,\hat{j}_m}}{\hat{\xi}_{t,i_m}^2}$. □

Example 3. Both ξ and A are not known explicitly; moreover, the ratios $\beta_{ij} = a_{ij}/p_{ij}$ are known for a certain transition matrix P with $p_{ij} \neq 0$ whenever $a_{ij} \neq 0$, and ξ is the unique invariant distribution of the Markov chain associated with P. While P is not explicitly known, it is assumed that a simulator is available that can generate transitions according to P.

To estimate R, we first express it as

$$R = \sum_{i,l,j} (\beta_{il}\beta_{jl}) \cdot \left(\xi_i p_{il} \cdot \frac{p_{jl}\xi_j}{\xi_l} \right) \cdot \phi(i)\phi(j)'.$$

Noticing that $\frac{p_{jl}\xi_j}{\xi_l}$ equals the steady-state conditional probability $P(X_{t-1} = j \mid X_t = l)$ for the Markov chain X_t, we thus generate a sequence of pairs of indices (i_t, j_t) as follows. Let $(i_0, i_1, \ldots)$ be a trajectory of the Markov chain. At $i_{t+1} = l$, we generate, using the uniform distribution, one sample (j, l) from the set of past transitions to l, $\{(i_{t_k - 1}, i_{t_k}) \mid i_{t_k} = l, t_k \leq t+1\}$, and we let $j_t = j$. (Indeed, this will also work if we simply let $j_t = i_{t_{k-1}}$ where t_k is the most recent time prior to $t+1$ that $i_{t_k} = l$.) It can be seen that the conditional probability of j_t given i_{t+1} converges asymptotically to $\frac{p_{j_t i_{t+1}} \xi_{j_t}}{\xi_{i_{t+1}}}$. We then define R_t by

$$R_t = \frac{1}{t+1} \sum_{m=0}^{t} (\beta_{i_m i_{m+1}} \beta_{j_m i_{m+1}}) \cdot \phi(i_m)\phi(j_m)',$$

and we approximate R by the symmetrized matrix $(R_t + R_t')/2$.

If the Markov chain is reversible, i.e., $\xi_j p_{jl} = \xi_l p_{lj}$ for all j, l, then the method can be substantially simplified. We can omit the procedure of generating j_t and simply set $j_m = i_{m+2}$ in R_t, because if we do so, the proper weight for the sample is $\frac{\xi_{j_m} p_{j_m i_{m+1}}}{\xi_{i_{m+1}} p_{i_{m+1} j_m}} = 1$. □

Example 4. The weight vector ξ is known explicitly, but A is not; moreover, the ratios $\beta_{ij} = a_{ij}/p_{ij}$ are known for a certain transition matrix P with $p_{ij} \neq 0$ whenever $a_{ij} \neq 0$. Here, ξ need not be the invariant distribution of P.

We can deal with this case by combining partially the schemes in Examples 2 and 3. We express R and generate a sequence of pairs of indices (i_t, j_t) as in Example 3. We keep tracking the empirical distribution κ_t of the sequence i_t up to time t, to approximate the invariant distribution of P. We weight samples properly to define R_t:

$$R_t = \frac{1}{t+1} \sum_{m=0}^{t} (\beta_{i_m i_{m+1}} \beta_{j_m i_{m+1}}) \cdot \left(\frac{\xi_{i_m} \xi_{j_m}}{\xi_{i_{m+1}}} \cdot \frac{\kappa_{t,i_{m+1}}}{\kappa_{t,i_m} \kappa_{t,j_m}} \right) \cdot \phi(i_m) \phi(j_m)',$$

and we approximate R by the symmetrized matrix $(R_t + R_t')/2$.

If the Markov chain associated with P is reversible, then there is simplification, similar to that in Example 3. We simply set $j_t = i_{t+2}$ and

$$R_t = \frac{1}{t+1} \sum_{m=0}^{t} (\beta_{i_m i_{m+1}} \beta_{i_{m+2} i_{m+1}}) \cdot \left(\frac{\xi_{i_m} \xi_{i_{m+2}}}{\xi_{i_{m+1}}} \cdot \frac{\kappa_{t,i_{m+1}}}{\kappa_{t,i_m} \kappa_{t,i_{m+2}}} \right) \cdot \phi(i_m) \phi(i_{m+2})',$$

because the extra term needed for weighting the sample properly is $\frac{\kappa_{t,j_m} p_{j_m i_{m+1}}}{\kappa_{t,i_{m+1}} p_{i_{m+1} j_m}}$, which converges to 1 as $m \to \infty$. □

A main source of difficulty in the estimation of R in MDP, as Examples 3 and 4 illustrate, is the unknown matrix A and the need of samples of "backward" transitions from a common state/index. Simulating backward transitions according to the steady-state conditional distribution is in general not easy. Consistently, as Example 1 illustrates, the estimation of R is quite simple when backward transitions can be easily generated, such as when A is known. A second source of difficulty in the estimation of R, as Examples 2-4 illustrate, is the memory demand. In particular, in order to either generate backward transitions or to weight samples properly, we must keep track of the past history of the simulation (except in the case of Example 3 and a reversible Markov chain).

Another drawback of the procedures given in Examples 1-4 is that they do not adapt easily to the case where A itself is a summation of infinitely many matrices, as in TD(λ) with $\lambda > 0$.

3 Applications

We consider two applications of Theorems 1 and 2. The first one is cost function approximation in MDP with TD-type methods. This includes single policy evaluation with discounted and undiscounted cost criteria, as well as the optimal cost approximation for optimal stopping problems. The second application is approximately solving large general systems of linear equations. We also illustrate with figures various issues discussed in Section 2.2 on the comparison of the bounds.

3.1 Cost Function Approximation for MDP

For policy evaluation in MDP, x^* is the cost function of the policy to be evaluated. Let P be the transition matrix of the Markov chain induced by the policy. The original linear equation that we want to solve is the Bellman equation, or optimality equation, satisfied by x^*. It takes the form

$$x^* = g + \alpha P x^*,$$

where g is the per-stage cost vector, and $\alpha \in [0, 1]$ is the discount factor: $\alpha \in [0, 1)$ corresponds to the discounted cost criterion, while $\alpha = 1$ corresponds to either the total cost criterion or the average cost criterion (in the latter case g is the per-stage cost minus the average cost). For simplicity of discussion, we assume that the Markov chain is irreducible.

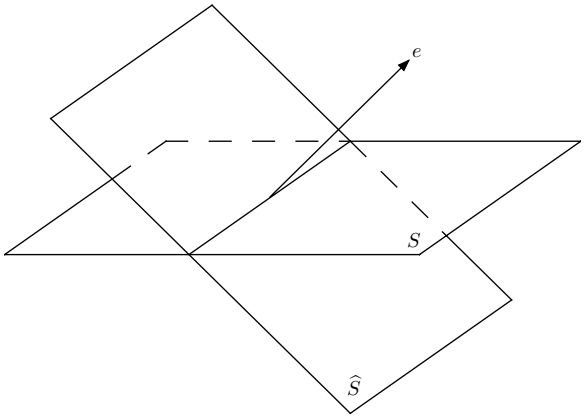


Figure 3: Illustration of $\widehat{S}$, the orthogonal complement of e in $S \oplus e$, i.e., $\widehat{S} = (S \oplus e) \cap e^{\perp}$.

With the TD(λ) method, we solve a projected form of the multistep Bellman equation

$$x = \Pi b + \Pi A x,$$

where the matrix A and the vector b are defined for a pair of values (α, λ) by

$$A = P^{(\alpha,\lambda)} \stackrel{def}{=} (1-\lambda)\sum_{l=0}^{\infty} \lambda^l (\alpha P)^{l+1}, \qquad b = \sum_{l=0}^{\infty} \lambda^l (\alpha P)^l g,$$

respectively, with either $\alpha \in [0,1), \lambda \in [0,1]$, or $\alpha = 1, \lambda \in [0,1)$. Notice that the case $\lambda = 0$ corresponds to $A = \alpha P, b = g$.

We note that for TD(λ) with $\lambda > 0$, we do not yet have an efficient simulation-based method for estimating the bound of Theorem 2; we have calculated the bound using common matrix algebra, and we plot it just for comparison.

Discounted Problems

Consider the discounted case: $\alpha < 1$. For $\lambda \in [0,1]$, with ξ being the invariant distribution of the Markov chain, the modulus of contraction of $P^{(\alpha,\lambda)}$ with respect to $\|\cdot\|_\xi$ is

$$\|P^{(\alpha,\lambda)}\|_\xi = \frac{(1-\lambda)\alpha}{1-\lambda\alpha}.$$

Let e denote the constant vector of all ones. Like P, the matrix $P^{(\alpha,\lambda)}$ has e as an eigenvector associated with the dominant eigenvalue $\frac{(1-\lambda)\alpha}{1-\lambda\alpha}$.

If the approximation subspace S contains or nearly contains e, the bound of Theorem 1 can degrade to the worst case error bound given by (2), as remarked in Section 2.2. In such a case, in order to have a sharper bound for the approximation of Πx^*, we can estimate separately the projection of x^* on e and the projection of x^* on another subspace $\widehat{S} = (S \oplus e) \cap e^{\perp}$, which is the orthogonal complement of e in $S \oplus e$ (see Figure 3), and redefine $\bar{x}$ as the sum of the two estimates. When the first projection can be estimated with no bias, the error bound for the second projection carries over to the combined estimate $\bar{x}$. This is true generally, not only for e, but for any eigenspace of P replacing e, as discussed in Section 2.2, Prop. 2 and Remark 4. In the case here, with ξ being the invariant distribution of the Markov chain, the projection of x^* on e can be calculated asymptotically exactly through simulation. It can be seen that the projection of x^* on e equals

$$\xi' x^* = \xi' b + \xi' P^{(\alpha,\lambda)} x^* = \xi' b + \frac{(1-\lambda)\alpha}{1-\lambda\alpha}\xi' x^*, \quad \Rightarrow \quad \xi' x^* = \frac{1-\lambda\alpha}{1-\alpha}\xi' b.$$

In addition, basis vectors of $\widehat{S}$ can also be generated from Φ by using simulation (we estimate the "mean feature," $\xi'\Phi$, and subtract it from the rows of Φ; see e.g., [Kon02]), along with the approximation of the matrices B and M and without incurring much computation overhead. Figure 4 illustrates the error bounds, and shows how the use of $\widehat{S}$ may improve them. It can be observed that the bound of Theorem 2 has consistently performed best, as indicated by the analysis.

Figure 5 compares the bounds for the case where the projection norm is the standard unweighted Euclidean norm. The standard bounds and the bound of Theorem 1 need the value $\|A\|$, while the bound of Theorem 2 does not. For comparison of these bounds, we compute $\|P\|$ using the knowledge of P, bound $\|A\|$ by $\frac{(1-\lambda)\|\alpha P\|}{1-\lambda\|\alpha P\|}$, and plug the latter in the standard bounds and the bound of Theorem 1. The value $\|\alpha P\|$, which corresponds to $\|A\|$ for $\lambda = 0$, is shown in the titles of Figure 5. With the norm being different from $\|\cdot\|_\xi$, the mapping ΠA is not necessarily a contraction for small values of λ, even though in this example it is.

Note that the availability of computable error bounds for non-contraction mappings facilitates the design of policy evaluation algorithms with improved exploration. In particular, we can use the LSTD algorithm [Boy99] to evaluate the cost or the Q-factor of a policy using special sampling methods that enhance exploration, and use the bound of Theorem 1 to estimate the corresponding amplification ratio.[4] Alternatively, we may use the bound of Theorem 2 in conjunction with TD(0)-type algorithms. Examples 3 and 4 show how to estimate the matrix R in cases where the projection norm is determined by an exploration policy, and where the projection norm is given explicitly with the desirable weights, respectively.

Average Cost and Stochastic Shortest Path (SSP) Problems

In the average cost case (similarly for SSP), x^* is the differential cost or bias vector and it is orthogonal to e. Let us assume that S is orthogonal to e, to simplify the discussion. Let ξ be the invariant distribution of the Markov chain. The error bound corresponding to the bound (3), as given by Tsitsiklis and Van Roy [TV99a], is

$$\|x^* - \bar{x}\|_\xi \leq \frac{1}{\sqrt{1-\alpha_\lambda^2}}\|x^* - \Pi x^*\|_\xi,$$

where $\alpha_\lambda < 1$ and $\alpha_\lambda \to 0$ as $\lambda \to 1$. Here, α_λ can be viewed as the modulus of contraction of some mapping that is a damped version of ΠA, while $\alpha_\lambda \to 0$ reflects the fact that the matrix ΠA converges to the zero matrix (as A converges to $e\xi'$) as $\lambda \to 1$. Note that the factor in the bound converges to 1, as $\lambda \to 1$. This bound is qualitative, as usually the value of α_λ is unknown.

Figure 6 shows the bounds of Theorems 1 and 2. Notice that as $\lambda \to 1$, the bound of Theorem 1 converges to $\sqrt{2}$ instead of 1. This is due to the over-relaxation in the analysis for the case where ΠA is near zero, as remarked in Section 2.2. Notice also in Figure 6(b) that the bound of Theorem 1 is affected by the relation of S to the eigenspace of A associated with eigenvalues that are close to 1, similar to the discounted case. By contrast, the bound of Theorem 2 performs well.

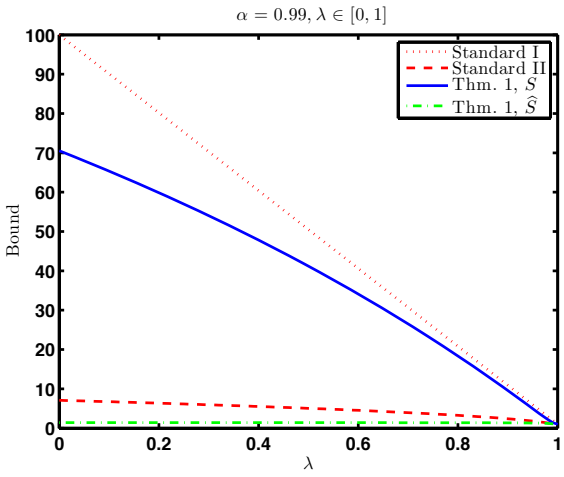
Optimal Stopping Problems

In optimal stopping problems, we have an uncontrolled Markov chain with transition matrix P, and we seek an optimal policy to stop the process so that we minimize the expected total (discounted or undiscounted) cost. With x^* being the optimal cost function, the Bellman equation is

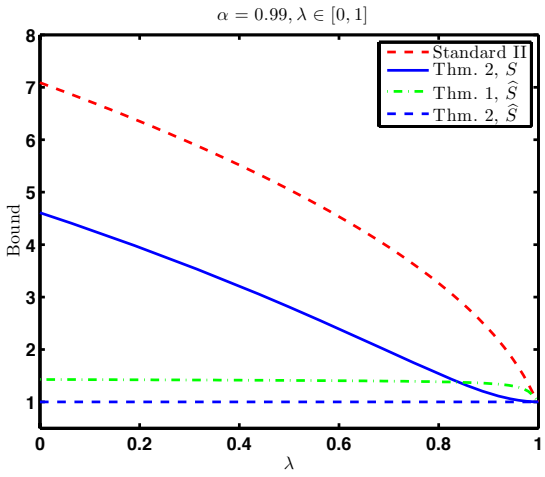
$$x^* = g + \alpha P \min\{c, x^*\},$$

where g is the vector of one-stage cost associated with continuation and c is the vector of one-stage cost associated with stopping. This is a nonlinear equation.

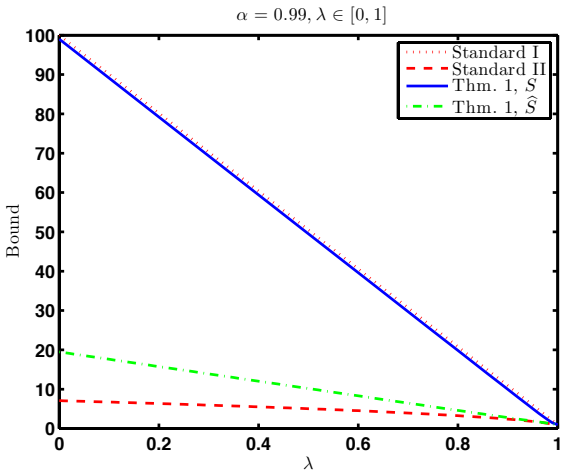
[4] When ΠA is not necessarily a contraction, a bound on $\|A\|_\xi$ is needed to apply Theorem 1. There are also algorithms involving exploration and maintaining the contraction property of ΠA, for which we refer to our paper [BY08].



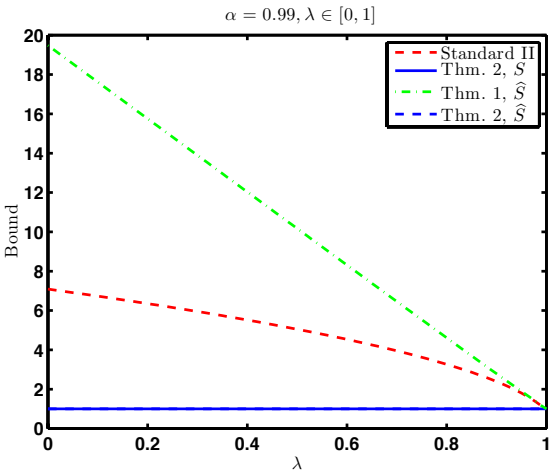
(a) Standard bounds vs. Theorem 1



(b) Standard bounds vs. Theorems 1 & 2 [detail of lower portion of (a)]

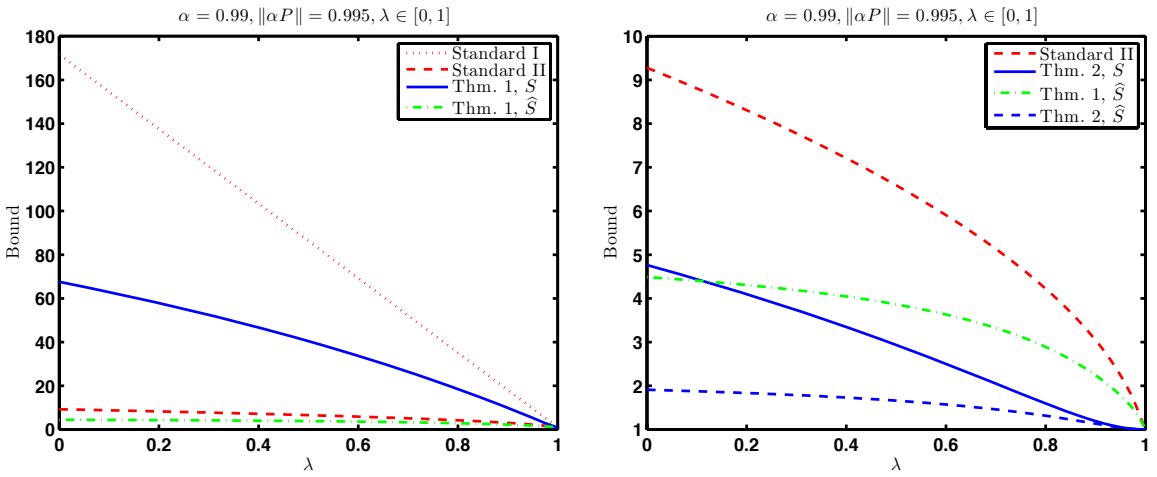


(c) Standard bounds vs. Theorem 1



(d) Standard bounds vs. Theorems 1 & 2 [detail of lower portion of (a)]

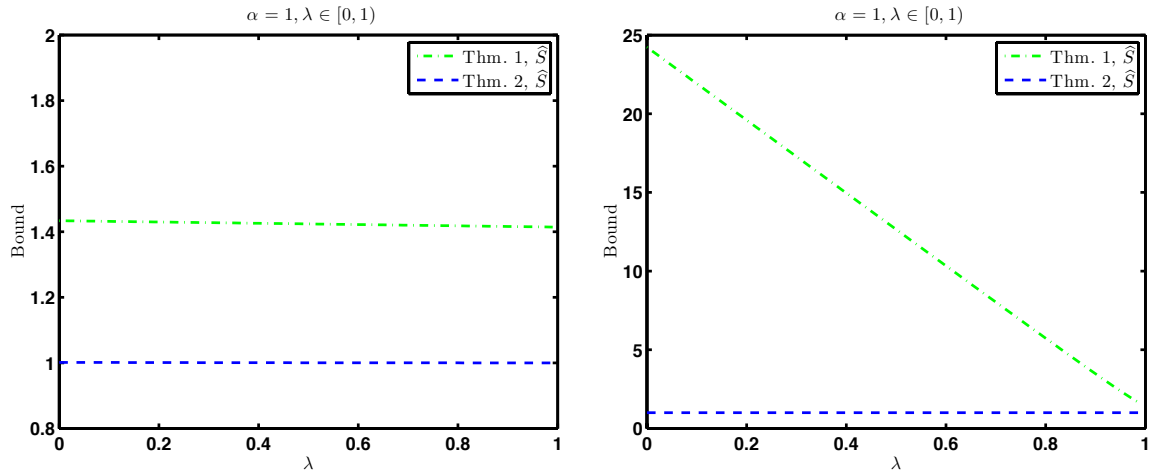
Figure 4: Comparison of error bounds as functions of λ for two discounted problems with randomly generated Markov chains. The dimension parameters are $n = 200, k = 50$, and the weights ξ in the projection norm is the invariant distribution. Standard I and II refer to the worst case bounds (2) and (3), respectively. The Markov chain is the same in (a) and (b), and in (c) and (d). In (c) and (d), the Markov chain has a "noisy" block structure with two blocks, thus P has a relatively large subdominant eigenvalue; S is chosen to contain e and a vector close to an eigenvector associated with that subdominant eigenvalue. The subspace $\widehat{S}$ is derived from S by orthogonalization, as shown in Figure 3.



(a) Standard bounds vs. Theorem 1

(b) Standard bounds vs. Theorems 1 & 2 [detail of lower portion of (a)]

Figure 5: Comparison of error bounds for discounted problems. The setup is the same as that for Figure 4, except that the projection norm is the standard Euclidean norm. The Markov chain has a "noisy" block structure. The subspace S is chosen randomly.



(a) Theorem 1 vs. Theorem 2

(b) Theorem 1 vs. Theorem 2

Figure 6: Comparison of error bounds for average cost problems with randomly generated Markov chains. The setup is the same as that for Figure 4. In (b), the Markov chain has a "noisy" block structure, and S is chosen as in Figure 4(c).

Let ξ be the invariant distribution of the Markov chain. Algorithms analogous to TD(0) [TV99b, CV06, YB06] solve the projected Bellman equation, which is also nonlinear,

$$x = \Pi g + \alpha \Pi P \min\{c, x\}.$$

There are error bounds analogous to the bound (3) and based on the contraction property of the mapping $\Pi P \min\{c, \cdot\}$ [TV99b, Van07].

To apply our error bounds, we shall form a linear equation based on the approximating solution $\bar{x}$, which satisfies

$$\bar{x} = \Pi g + \alpha \Pi P \min\{c, \bar{x}\}. \tag{30}$$

Let $I_{\bar{x}}$ be an $n \times n$ diagonal matrix with its i-th diagonal entry defined by

$$I_{\bar{x},ii} = \begin{cases} 1 & \bar{x}_i < c_i, \\ 0 & \text{otherwise.} \end{cases}$$

We consider the linear equation

$$x = g + \alpha P(I - I_{\bar{x}})c + \alpha P I_{\bar{x}} x, \tag{31}$$

and we view Eq. (30) as its projected form, i.e., we consider the following projected equation equivalent to Eq. (30):

$$x = \Pi\big(g + \alpha P(I - I_{\bar{x}})c\big) + \alpha \Pi P I_{\bar{x}} x.$$

Let $\hat{x}$ be the solution of Eq. (31). If $\bar{x} = x^*$, then $\hat{x} = \bar{x} = x^*$, so the difference $\hat{x} - \bar{x}$ provides some information about the approximation quality $x^* - \bar{x}$. We can apply the new error bounds with $A = \alpha P I_{\bar{x}}$ to bound $\hat{x} - \bar{x}$, once $\bar{x}$ is computed, and consequently the matrices and vectors in Eq. (31) are available. The matrices in the bounds can be estimated similar to those in [YB06]. Thus the new error bounds can provide supplementary information about the approximation quality, in addition to the error bounds based on the contraction property [TV99b, Van07].

3.2 Large General Systems of Linear Equations

For solving large general systems of linear equations using the projected equation approach [BY08], the bound of Theorem 2 can be computed in a straightforward way (except in the case of TD(λ) with $\lambda > 0$), as shown in Examples 1 and 2. Theorem 2 is not only much sharper than Theorem 1 for this case, but also more convenient, because it does not require the knowledge of $\|A\|_\xi$. Note that we can write linear equations of the form $Lx = q$ as $x = Ax + b$, with $A = I + cL$ and $b = -cq$ for any scalar c, and we can choose c to optimize the corresponding error bound.

4 Related Results

In this section, we will first address the two over-relaxations of the bound of Theorem 1, as discussed in Section 2.2, and derive improved error bounds (Props. 3 and 4). These bounds are qualitative and not easy to compute using data, however. We will then derive an analogous computable error bound (Prop. 5) for an alternative approximation approach, namely, the equation error minimization method. Our line of analysis is similar to the one of Section 2. In particular, Props. 4 and 5 require applications of Lemma 2 with different matrices, just like Theorems 1 and 2.

4.1 Two Additional Qualitative Error Bounds for Projected Equations

Error Bound Relating to Eigenspaces of A

We mentioned in Section 2.2 that when the approximation subspace S nearly contains eigenvectors of A corresponding to eigenvalues that are close to 1, the bound of Theorem 1 over-relaxes. This does

not imply that the actual approximation error $x^* - \bar{x}$ is big. On the contrary, including eigenspaces of A in S does not degrade the approximation quality in general: this can be seen from the analysis of Theorem 2 and can also be observed from the behavior of the bound of Theorem 2 as illustrated by the figures in Section 3. Our aim here is to improve over the bound of Theorem 1 qualitatively to capture the relation between the eigenspace of A and the approximation quality.

First, note that when $x^* - \Pi x^*$ is an eigenvector of A corresponding to some real eigenvalue c, we have $\bar{x} = \Pi x^*$, that is, the solution of the projected equation $\bar{x}$ coincides with the projection of x^*. This is because Πx^* satisfies the projected equation: from $\Pi A(x^* - \Pi x^*) = c\Pi(x^* - \Pi x^*) = 0$, we have

$$\Pi x^* = \Pi(Ax^* + b) = \Pi A(x^* - \Pi x^*) + \Pi A \Pi x^* + \Pi b = \Pi A \Pi x^* + \Pi b.$$

Similarly, when $x^* - \Pi x^*$ is close to such an eigenspace of A, $\bar{x}$ and Πx^* are also close to each other. We now make the analysis more precise by using the following fact: since $\Pi(x^* - \Pi x^*) = 0$, we have for any scalar c,

$$\Pi A(x^* - \Pi x^*) = \Pi(A - cI)(x^* - \Pi x^*). \tag{32}$$

We derive a bound in a form analogous to Theorem 1 by varying the choice of c. In particular, let us define a function $g : \Re^n \to \Re$ by

$$g(z) = \min_{c \in \Re} \|Az - cz\|_\xi .$$

Note that g is positively homogeneuous. Note also that g has the property $g(z) \leq \|A\|_\xi \|z\|_\xi$ (too see this, choose $c = 0$), and g vanishes at eigenvectors of A corresponding to real eigenvalues (to see this, choose c to be that eigenvalue). It can be seen that the optimal c^* is $c^* =< Az, z >_\xi / \|z\|_\xi^2$, so g can be expressed analytically as

$$g(z)^2 = \|Az\|_\xi^2 - c^* < Az, z >_\xi = \|Az\|_\xi^2 - \frac{< Az, z >_\xi^2}{\|z\|_\xi^2}. \tag{33}$$

(In the above, we define $0/0$ to be 0.) The bound we obtain is an improvement over the one of Theorem 1 and can be stated as follows.

Proposition 3. *The approximation error $x^* - \bar{x}$ satisfies*

$$\|x^* - \bar{x}\|_\xi \ \leq \ \sqrt{1 + \sigma(G_1)\, g\big(\tfrac{x^* - \Pi x^*}{\|x^* - \Pi x^*\|_\xi}\big)^2}\ \|x^* - \Pi x^*\|_\xi, \tag{34}$$

where G_1 is given by Eq. (10), and g is the positive homogeneous function given by Eq. (33) and has the property $g(z) = 0$ for any eigenvector z of A that corresponds to a real eigenvalue.

Proof. Let $y = x^* - \Pi x^*$. Using $\Pi = \Phi(\Phi' \Xi \Phi)^{-1} \Phi' \Xi$, we write Eq. (32) equivalently as

$$\Phi' \Xi A y = \Phi' \Xi (A - cI) y$$

for any scalar c. We can exploit this relation in the proof of Theorem 1: with $C = FB^{-1}$, we have [cf. the left hand side of Eq. (21)]

$$\Phi C \Phi' \Xi A y = \Phi C \Phi' \Xi (A - cI) y, \qquad \forall c \in \Re,$$

which implies that we can replace Eq. (21) by the following inequality

$$\|\Phi C \Phi' \Xi A y\|_\xi \leq \sqrt{\sigma(G_1)} \min_{c \in \Re} \|Ay - cy\|_\xi = \sqrt{\sigma(G_1)}\, g(y), \tag{35}$$

where G_1 is as given in Theorem 1. Since g is a positive homogeneous function, $g(z)$ can be expressed as $g(z) = g\big(\frac{z}{\|z\|_\xi}\big)\|z\|_\xi$, and the inequality (35) is equivalent to

$$\|\Phi C \Phi' \Xi A y\|_\xi \leq \sqrt{\sigma(G_1)}\, g\big(\tfrac{y}{\|y\|_\xi}\big)\|y\|_\xi,$$

which using the proof of Theorem 1, implies Eq. (34). □

Error Bound in a Decomposed Form

We now investigate how each component of the residual vector $x^* - \Pi x^*$ may affect the bias $\bar{x} - \Pi x^*$. The following proposition provides a decomposition of error bound. Each term in the bound can be estimated easily, even for TD(λ) with $\lambda > 0$, like in Theorem 1, while the analysis takes into account that $x^* - \Pi x^*$ is orthogonal to S, like in Theorem 2. Overall, the bound is still qualitative, as it is generally impractical to estimate all terms involved. On the other hand, if in a special case, some of the individual terms in the bound can be easily calculated, the result may be computationally useful.

Proposition 4. *Let $\widehat{S}_i, i = 1, \ldots, m$, be m mutually orthogonal subspaces such that*

$$x^* - \Pi x^* \in \oplus_{i=1}^m \widehat{S}_i.$$

Let Ψ_i be an $n \times \hat{k}_i$ matrix whose columns form a basis of $\widehat{S}_i$, $i = 1, \ldots, m$, respectively. Then

$$\|\bar{x} - \Pi x^*\|_\xi \leq \left(\sum_{i=1}^m \sqrt{\sigma\big(\widehat{G}_i\big)} \right) \|x^* - \Pi x^*\|_\xi , \tag{36}$$

where $\widehat{G}_i$ is the $\hat{k}_i \times \hat{k}_i$ matrix

$$\widehat{G}_i = \widehat{B}_i^{-1} \widehat{C}_i' B \widehat{C}_i$$

with

$$\widehat{C}_i = FB^{-1}(E_{i,1} - MB^{-1}E_{i,2}),$$

and

$$\widehat{B}_i = \Psi_i' \Xi \Psi_i, \qquad E_{i,1} = \Phi' \Xi A \Psi_i, \qquad E_{i,2} = \Phi' \Xi \Psi_i.$$

Furthermore,

$$\sigma\big(\widehat{G}_i\big) = \|(I - \Pi A)^{-1} \Pi A (I - \Pi) \widehat{\Pi}_i\|_\xi^2,$$

where $\widehat{\Pi}_i$ is the mapping of projection on $\widehat{S}_i$, so the bound (36) is invariant to the choice of basis vectors of $\widehat{S}_i, i = 1, \ldots, m$, and S (i.e., $\Psi_i, i = 1, \ldots, m$, and Φ).

Proof. Our line of analysis is the same as that for Theorem 2. Let us denote $y = x^* - \Pi x^*$ and $C = FB^{-1}$. The assumption implies that $y = \sum_{i=1}^m \widehat{\Pi}_i y$, so

$$\bar{x} - \Pi x^* = (I - \Pi A)^{-1} \Pi A (I - \Pi) y = \sum_{i=1}^m (I - \Pi A)^{-1} \Pi A (I - \Pi) \widehat{\Pi}_i y. \tag{37}$$

By Lemma 1 and the definition of $\widehat{\Pi}_i$,

$$(I - \Pi A)^{-1} \Pi A (I - \Pi) \widehat{\Pi}_i y = \Phi C \Phi' \Xi A (I - \Pi) \Psi_i \widehat{B}_i^{-1} \Psi_i' \Xi y.$$

Applying Lemma 2 to the matrix $\Phi C \Phi' \Xi A (I - \Pi) \Psi_i \widehat{B}_i^{-1} \Psi_i' \Xi$ with

$$H = \Phi C \Phi' \Xi A (I - \Pi) \Psi_i \widehat{B}_i^{-1}, \qquad D = \Psi_i' \Xi,$$

we have

$$\|(I - \Pi A)^{-1} \Pi A (I - \Pi) \widehat{\Pi}_i y\|_\xi \leq \sqrt{\sigma\big(G\big)}\, \|y\|_\xi, \tag{38}$$

where $G = (H' \Xi H)(D \Xi^{-1} D')$.

We now verify that G is the matrix $\widehat{G}_i$ given in the statement. We have

$$H = \Phi C \Phi' \Xi A (I - \Pi) \Psi_i \widehat{B}_i^{-1} = \Phi C (E_{i,1} - MB^{-1} E_{i,2}) \widehat{B}_i^{-1} = \Phi \widehat{C}_i \widehat{B}_i^{-1},$$

and so, with $D \Xi^{-1} D' = \widehat{B}_i$, we have

$$G = (H' \Xi H)(D \Xi^{-1} D') = \widehat{B}_i^{-1} \widehat{C}_i' B \widehat{C}_i,$$

which is the matrix $\widehat{G}_i$ given in the statement.

Combining Eqs. (37), (38), and the triangle inequality, we obtain the bound (36). The rest of proof is the same as that of Theorem 2. □

4.2 Error Bound for an Alternative Approximation Method

Our line of analysis in Section 2 can also be applied to obtain analogous data-dependent error bounds on the amplification/bias-to-distance ratio for a different approximation approach, which uses the solution of

$$\min_{x\in S} \|x - Ax - b\|_\xi^2 \tag{39}$$

as an approximation of x^*. We will make the assumption that $I - A$ is invertible, under which both x^* and the solution of (39) are unique.

Proposition 5. *Assume $I - A$ is invertible. Let $\tilde{x}$ be the solution of the minimization problem (39). Then,*

$$\|\tilde{x} - \Pi x^*\|_\xi \leq \sqrt{\sigma(\widetilde{G})}\, \|x^* - \Pi x^*\|_\xi, \tag{40}$$

where $\widetilde{G}$ is the $k \times k$ matrix

$$\widetilde{G} = B\widetilde{E}\widetilde{R}\widetilde{E} - I, \tag{41}$$

with

$$B = \Phi'\Xi\Phi, \qquad \widetilde{E} = (\Phi' L'\Xi L\Phi)^{-1}, \qquad \widetilde{R} = \Phi' L'\Xi L\Xi^{-1} L'\Xi L\Phi, \tag{42}$$

and $L = I - A$. Futhermore, the bound (40) is invariant to the choice of basis vectors of S (i.e., Φ).

Proof. First we establish an equality relation analogous to Eq. (4):

$$\tilde{x} - \Pi x^* = C(x^* - \Pi x^*) \tag{43}$$

where $C = \Phi(\Phi' L'\Xi L\Phi)^{-1}\Phi' L'\Xi L$. We then apply Lemma 2, taking into account that $\Pi(x^* - \Pi x^*) = 0$, similar to the analysis for Theorem 2.

The minimization problem (39) can be written as $\min_{r\in\Re^k} \|L\Phi r - b\|_\xi^2$. The optimality condition is

$$\Phi' L'\Xi(L\tilde{x} - b) = 0.$$

Since $Lx^* - b = 0$, we also have

$$\Phi' L'\Xi(L\Pi x^* - b) = \Phi' L'\Xi L(\Pi x^* - x^*).$$

Subtracting the last two equations, we have

$$\Phi' L'\Xi L(\tilde{x} - \Pi x^*) = \Phi' L'\Xi L(x^* - \Pi x^*).$$

Since L is invertible and Φ has full rank, $\Phi' L'\Xi L\Phi$ is invertible. Multiplying both sides of the above equation by $\Phi(\Phi' L'\Xi L\Phi)^{-1}$, and using the fact that $\tilde{x} - \Pi x^* = \Phi r \in S$ for some r, we obtain

$$\tilde{x} - \Pi x^* = \Phi(\Phi' L'\Xi L\Phi)^{-1}\Phi' L'\Xi L(x^* - \Pi x^*),$$

which is Eq. (43).

Since $\Pi(x^* - \Pi x^*) = 0$, Eq. (43) is further equivalent to

$$\tilde{x} - \Pi x^* = \Phi(\Phi' L'\Xi L\Phi)^{-1}\Phi' L'\Xi L(I - \Pi)(x^* - \Pi x^*). \tag{44}$$

Therefore,

$$\|\tilde{x} - \Pi x^*\|_\xi \leq \|\Phi(\Phi' L'\Xi L\Phi)^{-1}\Phi' L'\Xi L(I - \Pi)\|_\xi \|x^* - \Pi x^*\|_\xi.$$

Applying Lemma 2 to the matrix in the right hand side with

$$H = \Phi, \qquad D = (\Phi' L'\Xi L\Phi)^{-1}\Phi' L'\Xi L(I - \Pi) = \widetilde{E}\Phi' L'\Xi L(I - \Pi),$$

we have

$$\|\tilde{x} - \Pi x^*\|_\xi \leq \sqrt{\sigma(G)}\, \|x^* - \Pi x^*\|_\xi,$$

where $G = (H'\Xi H)(D\Xi^{-1}D') = B(D\Xi^{-1}D')$. Similar to the calculation in the proof of Theorem 2, it can be shown that

$$D\Xi^{-1}D' = \widetilde{E}\big(\widetilde{R} - \widetilde{M}B^{-1}\widetilde{M}'\big)\widetilde{E}',$$

where

$$\widetilde{R} = \Phi'L'\Xi L\Xi^{-1}L'\Xi L\Phi, \qquad \widetilde{M} = \Phi'L'\Xi L\Phi = \widetilde{E}^{-1}.$$

Thus,

$$G = B\widetilde{E}\widetilde{R}\widetilde{E} - I,$$

which is the matrix $\widetilde{G}$ given in Eq. (41).

Finally we prove that the bound is invariant to the choice of basis vectors of S. To see this, we write the matrix $\Phi(\Phi'L'\Xi L\Phi)^{-1}\Phi'L'\Xi L(I-\Pi)$ equivalently as the product of three matrices: $L^{-1}\cdot L\Phi(\Phi'L'\Xi L\Phi)^{-1}\Phi'L'\Xi\cdot L(I-\Pi)$. The first and the third matrices clearly do not depend on the choice of Φ, while the second matrix is the projection mapping on the subspace $L(S)$, hence it also does not depend on the choice of Φ. Therefore, the bound is invariant to the choice of Φ. □

Remark 6. Similar to the argument in the proof of Prop. 1, one can show that the bound of Prop. 5 is tight, in the sense that for any A and S, there exists a worst case choice of b for which the bound holds with equality.

5 Conclusion

We have considered the projected equation approximation approach, and we have presented new data-dependent computable error bounds that hold for both contraction and non-contraction mappings. Their applicability for non-contraction mappings is not only useful for approximating solutions of general linear equations, but is also useful in the context of MDP for designing exploration mechanisms. Furthermore, in the context of MDP, these bounds can be used in performance bounds for approximate policy iteration, such as the ones of [Mun03].

One potential use of our bounds is to suggest changes in the projected equation in order to reduce the amplification ratio. For example, extensive computational experience with TD(λ) methods suggests that the simulation noise tends to increase as λ increases, so there is motivation to use small values of λ as long as the amplification ratio is close to 1. Unfortunately, the bounds (2), (3) are too conservative to provide useful information about the amplification ratio, and our bounds can provide quantitative guidance as well as valuable insight in this regard. Furthermore, our bounds can be similarly used in the general non-contraction context, in conjunction with simulation-based TD(λ)-like algorithms that have been developed in our recent paper [BY08]. There may be other potential uses of our bounds, for example in suggesting changes to the choice of approximation subspace, thereby affecting both the baseline error and the amplification ratio, but this is a subject for future research.

Acknowledgment

Huizhen Yu is supported in part by Academy of Finland grant 118653 (ALGODAN) and by the IST Programme of the European Community, PASCAL Network of Excellence, IST-2002-506778. Dimitri Bertsekas is supported by NSF Grant ECCS-0801549.

References

[Ber07] D. P. Bertsekas, *Dynamic programming and optimal control*, third ed., vol. II, Athena Scientific, Belmont, MA, 2007.

[Boy99] J. A. Boyan, *Least-squares temporal difference learning*, Proc. The 16th Int. Conf. Machine Learning, 1999.

[BT96] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-dynamic programming*, Athena Scientific, Belmont, MA, 1996.

[BY08] D. P. Bertsekas and H. Yu, *Projected equation methods for approximate solution of large linear systems*, J. Computational and Applied Mathematics (2008), to appear.

[CV06] D. S. Choi and B. Van Roy, *A generalized Kalman filter for fixed point approximation and efficient temporal-difference learning*, Discrete Event Dyn. Syst. **16** (2006), no. 2, 207–239.

[Kon02] V. R. Konda, *Actor-critic algorithms*, Ph.D. thesis, MIT, Cambridge, MA, 2002.

[Mun03] R. Munos, *Error bounds for approximate policy iteration*, Proc. The 20th Int. Conf. Machine Learning, 2003.

[NB03] A. Nedić and D. P. Bertsekas, *Least squares policy evaluation algorithms with linear function approximation*, Discrete Event Dyn. Syst. **13** (2003), 79–110.

[SB98] R. S. Sutton and A. G. Barto, *Reinforcement learning*, MIT Press, Cambridge, MA, 1998.

[Sut88] R. S. Sutton, *Learning to predict by the methods of temporal differences*, Machine Learning **3** (1988), 9–44.

[TV97] J. N. Tsitsiklis and B. Van Roy, *An analysis of temporal-difference learning with function approximation*, IEEE Trans. Automat. Contr. **42** (1997), no. 5, 674–690.

[TV99a] ———, *Average cost temporal-difference learning*, Automatica **35** (1999), no. 11, 1799–1808.

[TV99b] ———, *Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing financial derivatives*, IEEE Trans. Automat. Contr. **44** (1999), 1840–1851.

[Van07] B. Van Roy, *On regression based stopping times*, 2007, manuscript.

[YB06] H. Yu and D. P. Bertsekas, *A least squares Q-learning algorithm for optimal stopping problems*, LIDS Tech. Report 2731, MIT, 2006.