MATHEMATICS OF OPERATIONS RESEARCH
Vol. 35, No. 2, May 2010, pp. 306–329
ISSN 0364-765X | EISSN 1526-5471 | 10 | 3502 | 0306

DOI 10.1287/moor.1100.0441
© 2010 INFORMS

Error Bounds for Approximations from Projected Linear Equations

Huizhen Yu
Department of Computer Science, University of Helsinki, FIN-00014 Helsinki, Finland, janey.yu@cs.helsinki.fi

Dimitri P. Bertsekas
Laboratory for Information and Decision Systems, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, dimitrib@mit.edu

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 Markov decision processes (MDP), one of our bounds is always sharper than the standard contraction-based bounds, and another one is often sharper. The former bound is also tight in a worst-case sense. Our bounds also apply to the noncontraction 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.

Key words: projected linear equations; function approximation; error bounds; dynamic programming; temporal difference methods; Galerkin methods
MSC2000 subject classification: Primary: 90C40, 41A45; secondary: 90C59
OR/MS subject classification: Primary: dynamic progamming/optimal control: Markov models; analysis of algorithms: suboptimality
History: Received August 9, 2008; revised August 3, 2009. Published online in *Articles in Advance* April 14, 2010.

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 $\|\cdot\|_\xi$, where $\xi \in \Re^n$ is a vector of positive components. We assume that x^* and $\bar{x}$ exist, and that the matrix $\mathrm{I} - \Pi A$ is invertible, so $\bar{x}$ 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 n-dimensional vector-matrix operations are practically impossible, while $k \ll n$. This approach is common in approximate dynamic programming (DP) for Markov decision processes (MDP) and has been central in much of the recent research on the subject (see, e.g., Sutton [14], Tsitsiklis and Van Roy [17], Bertsekas and Tsitsiklis [3], Sutton and Barto [15], Bertsekas [2]). Let us give the background of two important applications in this context. For 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., Sutton [14], Tsitsiklis and Van Roy [17], Bertsekas and Tsitsiklis [3], Sutton and Barto [15], Bertsekas [2]). More specifically, in this context, x^* is the cost vector of the policy to be evaluated; the original equation $x = Ax + b$ is the Bellman equation ($\lambda = 0$) or a multistep Bellman equation ($\lambda \in (0, 1]$); the weight vector ξ in the projection norm often corresponds to an invariant distribution of the Markov chain associated with that policy; and the subspace S is determined indirectly by certain numerical features characterizing the states of the Markov chain, without the need to store n-dimensional basis vectors. Another important application of the approximation approach in Equation (1), in the context of MDP, is in policy gradient methods of the actor-critic type (see e.g., Konda and Tsitsiklis [10], Konda [9]). There, one evaluates parametrized policies, and the original equation again corresponds to the Bellman equation or its multistep version. To estimate the gradient of the cost of a policy, the projected equation is solved and the solution $\bar{x}$ is used to approximate Πx^*, the projection of some solution of the original equation. The smaller the deviation of $\bar{x}$ from Πx^*, the smaller is the bias in the gradient estimation.

For the preceding MDP applications, often the model of the MDP that determines the values of A and b is either unavailable or too large to be represented explicitly, and only trajectories of the Markov chain that carry noisy information about the problem are observed or generated using a simulator. Then the projected equation approach is applied together with simulation and stochastic approximation techniques, which enable one to construct and solve the low-dimensional projected equation in a model-free manner. Applicability under an unknown model is another computational advantage of the projected equation approach and substantially expands its range of practical MDP applications.

For another important context, we note that the projected equation approach of Equation (1) belongs to the class of Galerkin methods and finds broad application in the approximate solution of linear operator equations; see, e.g., Krasnose'skii et al. [11]. For example, important finite element and other methods for solving partial differential equations belong to the Galerkin class. In our recent paper (Bertsekas and Yu [4]), 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, using the Monte Carlo simulation ideas that are central in approximate DP. While our work in Bertsekas and Yu [4] and the present work have been primarily motivated by the MDP context, they might find substantial application within the Galerkin approximation context.

The focus of the present paper is on analyzing the approximation error of the projected equation approach. The distance/error between x^* and $\bar{x}$ comprises two components (by the Pythagorean theorem):

$$\|x^* - \bar{x}\|_\xi = \sqrt{\|x^* - \Pi x^*\|_\xi^2 + \|\Pi x^* - \bar{x}\|_\xi^2}. \tag{2}$$

The first component $\|x^* - \Pi x^*\|_\xi$, which is the distance between x^* and Πx^*, the best approximation of x^* on the approximation subspace S with respect to $\|\cdot\|_\xi$, is the minimum error possible with any approximation methods given S. Our focus will be on bounding the second component $\|\Pi x^* - \bar{x}\|_\xi$, which is the bias in $\bar{x}$ relative to the best approximation Πx^*, due to solving the projected equation. In the two aforementioned application contexts, what adds to the complexity of error analysis is the need for bounds that can be practically evaluated when the dimension n is very large, as well as when the model (A and b) is only indirectly available through simulation. This is also an important characteristic that differentiates the analysis of the present paper from those in the classic framework of Galerkin methods.

In the MDP context, for the case where ΠA is 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., Bertsekas and Tsitsiklis [3], Tsitsiklis and Van Roy [17]) holds if $\|\Pi A\| = \alpha < 1$ with respect to some norm $\|\cdot\|$ and has the form

$$\|x^* - \bar{x}\| \le \frac{1}{1-\alpha}\|x^* - \Pi x^*\|. \tag{3}$$

The second bound (see, e.g., Tsitsiklis and Van Roy [18], Bertsekas [2]) holds in the 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 (2), and it is much sharper than the first bound:

$$\|x^* - \bar{x}\|_\xi \le \frac{1}{\sqrt{1-\alpha^2}}\|x^* - \Pi x^*\|_\xi. \tag{4}$$

The bounds (3), (4) are determined by the modulus of contraction α. In the MDP context, α is known or can be upper bounded by a known number, based on two parameters in the definition of ΠA, the discount factor of the MDP and the value of λ used in the TD algorithms. For example, for a discounted problem with discount factor 0.99 and TD(0), α can be bounded by 0.99, and an application of (4) yields $\|x^* - \bar{x}\|_\xi \le 7.1\|x^* - \Pi x^*\|_\xi$; similarly, with the discount factor being 0.999, $\|x^* - \bar{x}\|_\xi \le 22.4\|x^* - \Pi x^*\|_\xi$. Although easy to compute, these bounds tend to be conservative, particularly when the modulus of contraction of ΠA approaches 1 and the bounds correspondingly tend to infinity. These bounds provide a qualitative guarantee of the approximation quality of $\bar{x}$ for those approximation subspaces that are close to the set of solutions x^*. But because they do not use problem-dependent information, such as the relation between the subspace S and the matrix A, they do not fully reflect the nature of the error/bias of the projected equation approach (as also indicated by their reliance on the contraction property), and they cannot explain the causes of a large bias when the subspace S is reasonably good, or of a small bias when S is far away from the set of solutions x^*.

The case where ΠA is not a contraction mapping is relevant not only for approximating solutions of general linear equations, but also for MDP applications. For example, when evaluating a policy, it is often desirable to occasionally deviate from that policy, and to explore states and controls that do not frequently occur under that policy. Such exploration mechanisms result in a projected equation in which the weights in the projection norm no longer correspond to the invariant distribution of the Markov chain associated with the policy to be evaluated. As a result, ΠA need not be a contraction mapping.

The preceding contexts motivate us to develop error bounds of the general form

$$\|x^* - \bar{x}\|_\xi \le \mathscr{B}(A, \xi, S)\|x^* - \Pi x^*\|_\xi, \tag{5}$$

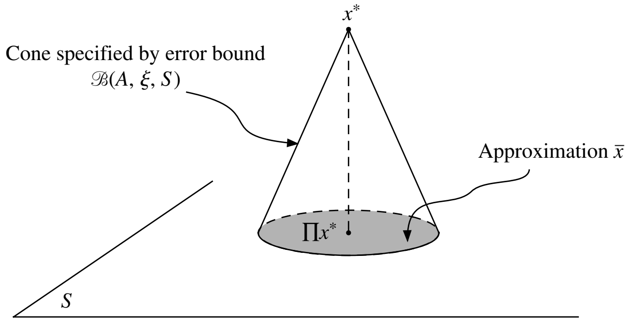


FIGURE 1. The relation between the error bound and $\bar{x}$: $\bar{x}$ lies in the intersection of S and a cone that originates from x^* and whose angle is specified by the error bound $\mathscr{B}(A, \xi, S)$ as $\cos^{-1}(1/(\mathscr{B}(A, \xi, S)))$.
Notes. The smaller $\mathscr{B}(A, \xi, S)$ is, the sharper the cone is. The smallest bound $\mathscr{B}(A, \xi, S) = 1$ implies $\bar{x} = \Pi x^*$.

where $\mathscr{B}(A, \xi, S)$ is a constant that depends on A, ξ, and S (but not on b), in contrast to the fixed error bounds (3), (4). The general error bounds (5) not only can be sharper than the standard bounds for the case, where ΠA is a contraction, but also apply when ΠA is not a contraction. Like the bounds (3), (4), 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 $\mathscr{B}(A, \xi, S)$ or equivalently $\sqrt{\mathscr{B}(A, \xi, S)^2 - 1}$ as an upper bound to the *amplification ratio* or the *"bias-to-distance" ratio*,

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

respectively. We note that these two ratios can be large and can be a significant cause for concern, as illustrated by examples given in Bertsekas [1] (see also Bertsekas and Tsitsiklis [3, Example 6.5, pp. 288–289]). Figure 1 illustrates the relation between the bound (5), x^* and $\bar{x}$.

We derive mathematical expressions for $\mathscr{B}(A, \xi, S)$ that involve the spectral radii of small-size matrices, which can be computed with low-dimensional operations, and simulation in the above application contexts, thus providing a "data/problem-dependent" error analysis, in contrast to the error bounds (3), (4). These expressions are derived by using the same line of analysis but with different degrees of concern for the ease of evaluating the expressions in practice. All our bounds are independent of the parametrization of the subspace S. In particular, we will derive two bounds:

$$\mathscr{B}(A, \xi, S) = \sqrt{1 + \sigma(G_1)\|\Pi A\|_\xi^2} \qquad \text{and} \qquad \mathscr{B}(A, \xi, S) = \sqrt{1 + \sigma(G_2)},$$

where G_1, G_2 are $k \times k$ matrices and $\sigma(\cdot)$ denotes the spectral radius of a matrix; see Theorems 2.1 and 2.2. The matrix in the first bound is easy to compute for all TD-type methods, and in fact it can be readily computed as a by-product of least-squares-based TD algorithms (e.g., Bradtke and Barto [6], Boyan [5], Nedić and Bertsekas [13], Bertsekas and Yu [4]). The second bound is sharper than the first. It is, in fact, tight for a worst-case choice of b; see Proposition 2.1 and Remark 2.3. Computing the matrix in the bound using simulation involves more complex schemes, which we will demonstrate for applications involving TD(0)-type methods. We will derive two additional bounds, one for characterizing the effect of a component of $x^* - \Pi x^*$ on the bias in $\bar{x}$, and the other for bounding a component of the approximation error $x^* - \bar{x}$; see Propositions 3.1 and 3.2. We will also use a similar approach to derive a bound of the form (5) for an alternative regression-based approximation method, the minimization of the equation error

$$\min_{x \in S} \|x - Ax - b\|_\xi^2, \tag{6}$$

under the additional assumption that $\mathrm{I} - A$ is invertible; see Proposition 3.3. This bound also involves small-size matrices that can be computed by simulation.

In connection with the Galerkin methodology, let us note that the approximation based on equation projection [cf. Equation (1)] is known as the Bubnov-Galerkin method, while the regression-based least-squares-error minimization [cf. Equation (6)] is a special case of the Petrov-Galerkin method; see, e.g., Krasnose'skii et al. [11]. Error analysis in the classic framework of Galerkin methods focuses on the asymptotic convergence and rate of convergence of $\bar{x}$ to x^* as the approximation subspace expands and $x^* - \Pi x^*$ diminishes. For this purpose, error bounds with analytical expressions suffice. Our error analysis departs from the classical analysis in that a fixed approximation subspace is considered, and instead of bounds with analytical expressions, bounds with practically computable or partially computable expressions are sought. To our knowledge, the form of our error bounds and the computational approach based on Monte Carlo simulation are new in the context of Galerkin methods.

We note also that the type of error bounds we develop in this paper are a priori error estimates, in the sense that they do not involve the equation error at $\bar{x}$ and can be obtained before solving the approximating problems for $\bar{x}$. A useful a posteriori error bound involving the equation error is

$$\|\bar{x}-x^*\|_\xi \le \|(\mathrm{I}-A)^{-1}\|_\xi \|\bar{x}-A\bar{x}-b\|_\xi,$$

which holds for any vector $\bar{x}$, assuming the invertibility of $\mathrm{I}-A$, as can be seen from the relation $\bar{x}-x^* = (\bar{x}-A\bar{x}-b)+(A\bar{x}+b)-(Ax^*+b)$, which implies $\bar{x}-x^* = (\mathrm{I}-A)^{-1}(\bar{x}-A\bar{x}-b)$. Note that this error bound makes use of the value of b and therefore can be sharper than ours, which hold for all values of b. However, in practice one must often solve a problem for a fixed A and a priori unknown multiple values of b, in which case a posteriori bounds that depend on b are less useful. Furthermore, while the equation error $\|\bar{x}-A\bar{x}-b\|_\xi$ can be estimated from simulation data after $\bar{x}$ is computed, it is generally difficult to compute or bound $\|(\mathrm{I}-A)^{-1}\|_\xi$ using simulation data. In the case where $\|A\|_\xi \le \alpha < 1$, i.e., A is a contraction mapping with respect to $\|\cdot\|_\xi$, one can bound $\|(\mathrm{I}-A)^{-1}\|_\xi$ by $1/(1-\alpha)$. But this leads to an overly conservative error estimate, like the bound (3), unless the equation error is rather small. It is still unclear whether one can use an approach similar to that of the present paper to sharpen the bound in this case. On the other hand, not using the a posteriori equation error and/or the vector b is a limitation of our a priori error estimates, which makes them at best tight in some worst-case sense. Combining them with a posteriori error estimates to obtain more powerful bounds is a subject for future research.

We present our main results in the next section and additional related results in §3. In §4, we address the application of the new error bounds to approximate policy evaluation in MDP and to the general problem of approximate solution of large systems of linear equations. In §5, we address methods for the estimation of the matrices in the bounds.

2. Main results.

We first introduce the main theorems and explain the underlying ideas and then give the proofs in §2.1 and compare the bounds in §2.2. 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 $x = \Pi(Ax+b)$.

The starting point for our bounding approach is the following equality, which relates the error/bias with the baseline error, and holds without contraction assumptions:

$$x^*-\bar{x} = (\mathrm{I}-\Pi A)^{-1}(x^*-\Pi x^*). \tag{7}$$

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

$$\Pi x^*-\bar{x} = \Pi A(x^*-\bar{x}) \quad \Rightarrow \quad (\Pi x^*-x^*)+(x^*-\bar{x}) = \Pi A(x^*-\bar{x}) \quad \Rightarrow \quad (7).$$

We express $(\mathrm{I}-\Pi A)^{-1}$ in the form $(\mathrm{I}-\Pi A)^{-1} = I+(\mathrm{I}-\Pi A)^{-1}\Pi A$, and correspondingly, the error as

$$x^*-\bar{x} = (x^*-\Pi x^*)+(\mathrm{I}-\Pi A)^{-1}\Pi A(x^*-\Pi x^*). \tag{8}$$

We aim at bounding the second term $(\mathrm{I}-\Pi A)^{-1}\Pi A(x^*-\Pi x^*)$ *directly*; this term is the bias in $\bar{x}$ relative to Πx^*:

$$\Pi x^*-\bar{x} = (\mathrm{I}-\Pi A)^{-1}\Pi A(x^*-\Pi x^*). \tag{9}$$

In doing so, we will obtain bounds that not only can be much sharper than the bounds (3) and (4) for the contraction case but also apply to the noncontraction case.

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

$$B = \Phi'\Xi\Phi, \qquad M = \Phi'\Xi A\Phi, \qquad F = (\mathrm{I}-B^{-1}M)^{-1}. \tag{10}$$

We will show later that the inverse in the definition of F exists (Lemma 2.1). 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.

We note that alternative matrix expressions in the subsequent bounds could be obtained by using a result of Horn and Johnson [8, Theorem 1.3.20], which states that $\sigma(U_1U_2) = \sigma(U_2U_1)$ for any $n\times m$ matrix U_1 and $m\times n$ matrix U_2.

Theorem 2.1. *The approximation error $x^* - \bar{x}$ satisfies*

$$\|x^* - \bar{x}\|_\xi \le \sqrt{1 + \sigma(G_1)\|\Pi A\|_\xi^2}\, \|x^* - \Pi x^*\|_\xi, \tag{11}$$

where G_1 is the $k \times k$ matrix

$$G_1 = B^{-1}F'BF. \tag{12}$$

Furthermore, $\sigma(G_1) = \|(\mathrm{I} - \Pi A)^{-1}\Pi\|_\xi^2$, so the bound (11) is invariant with respect to the choice of basis vectors of S (i.e., Φ).

The idea of the proof of Theorem 2.1 is to combine Equation (8) with the bound

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

and to show that $\|(\mathrm{I} - \Pi A)^{-1}\Pi\|_\xi^2 = \sigma(G_1)$. An important fact is that G_1 can be obtained by simulation, using low-dimensional calculations: The matrices B and M define the solution $\bar{x}$, and estimating them is indeed the basic procedure in several least-squares-based TD algorithms for computing $\bar{x}$ (e.g., Bradtke and Barto [6], Boyan [5], Nedić and Bertsekas [13], Bertsekas and Yu [4]), so with these algorithms, the bound can be obtained together with the approximate solution $\bar{x}$ without extra computation overhead.

While the bound of Theorem 2.1 can be conveniently computed, it is less sharp than the bound of the subsequent Theorem 2.2, and under certain circumstances less sharp than the contraction-based bound (4). In Theorem 2.1, $\|\Pi A\|_\xi$ is needed. It can be bounded by $\|A\|_\xi$, which in turn can be bounded by 1 when A is nonexpansive and by a known number for certain MDP applications involving noncontraction mappings (see §4). Generally, however, in the noncontraction case it might be hard to compute or estimate $\|A\|_\xi$. In Theorem 2.2, $\|A\|_\xi$ is no longer needed; A is absorbed into the matrix to be estimated. Furthermore, Theorem 2.2 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 2.2 involves a $k \times k$ matrix R (defined below) in addition to B and M, which might not be straightforward to estimate in some cases, as will be commented later.

Theorem 2.2. *The approximation error $x^* - \bar{x}$ satisfies*

$$\|x^* - \bar{x}\|_\xi \le \sqrt{1 + \sigma(G_2)}\|x^* - \Pi x^*\|_\xi, \tag{13}$$

where G_2 is the $k \times k$ matrix

$$G_2 = B^{-1}F'BFB^{-1}(R - MB^{-1}M'), \tag{14}$$

and R is the $k \times k$ matrix $R = \Phi'\Xi A \Xi^{-1} A' \Xi \Phi$. Furthermore, $\sigma(G_2) = \|(\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)\|_\xi^2$, so the bound (13) is invariant with respect to the choice of basis vectors of S (i.e., Φ).

The idea in deriving Theorem 2.2 is to combine Equation (8) with the bound

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

and to show that $\|(\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)\|_\xi^2 = \sigma(G_2)$. Incorporating the matrix $\mathrm{I} - \Pi$ in the definition of G_2 is crucial for improving the bound of Theorem 2.1. Indeed, the bound of Theorem 2.2 can be shown to be 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 (Proposition 2.1 and Remark 2.3 in §2.2).

Estimating the matrix R, although not always as straightforward as estimating B and M, is still possible in 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 multistep projected Bellman equations solved by TD(λ) with $\lambda > 0$. We will address these issues in §5.

Thus, Theorem 2.2 supersedes Theorem 2.1 in terms of sharpness of bounds, but the ease of evaluating the bound makes Theorem 2.1 still useful in practice. In §2.2 we will compare the bounds and discuss weaknesses in the bound of Theorem 2.1 as well as ways to improve it. The proofs given next contain the line of analysis that will also be useful in deriving additional bounds in §3.

2.1. Proofs of theorems. We shall need two technical lemmas. The first lemma introduces an expression of the matrix $(\mathrm{I}-\Pi A)^{-1}$ 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.

Recall that Φ is an $n\times k$ matrix whose columns form a basis of S. As a linear mapping, Φ defines a one-to-one correspondence between $\Re^k$ and S. Its inverse mapping is the unique linear transformation from S to $\Re^k$, denoted Φ^{-1}, which has the property $\Phi^{-1}(\Phi r)=r$ for all $r\in\Re^k$. (The inverse mapping Φ^{-1} has a representation as a $k\times n$ matrix, and because by the definition (10) of B, we have $(B^{-1}\Phi'\Xi)(\Phi r)=r$ for all $r\in\Re^k$, a matrix representation of Φ^{-1} is $B^{-1}\Phi'\Xi$.)

Lemma 2.1. *The matrix* $\mathrm{I}-\Pi A$ *is invertible if and only if the inverse* $(\mathrm{I}-B^{-1}M)^{-1}$ *defining* F *exists. When* $\mathrm{I}-\Pi A$ *is invertible,* $(\mathrm{I}-\Pi A)^{-1}$ *maps* S *onto* S, *and furthermore,*

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

Proof. For a mapping T whose domain contains S, let $T|_S$ denote T restricted to S. We have that $\mathrm{I}-\Pi A$ maps S into S. The matrix $\mathrm{I}-\Pi A$ is invertible if and only if there exists no $z\in\Re^n$, $z\neq 0$ with $\Pi Az=z$. Since such vectors z, when they exist, lie necessarily in S, we see that $\mathrm{I}-\Pi A$ is invertible if only if (i) holds, where

(i): $\mathrm{I}-\Pi A$ restricted to S is invertible, so $(\mathrm{I}-\Pi A)|_S$ defines a one-to-one mapping from S to S.

The statement in (i) is in turn equivalent to

(ii): The three-mapping composition

$$H=\Phi^{-1}\cdot(\mathrm{I}-\Pi A)|_S\cdot\Phi \tag{16}$$

defines a one-to-one mapping from $\Re^k$ to $\Re^k$, so H^{-1} exists.

Hence $\mathrm{I}-\Pi A$ is invertible if and only if H^{-1} exists. Using Equation (16) and $\Pi=\Phi B^{-1}\Phi'\Xi$, we have

$$H=\Phi^{-1}\cdot(\mathrm{I}-\Pi A)|_S\cdot\Phi=\mathrm{I}-\Phi^{-1}\cdot\Pi A\cdot\Phi=\mathrm{I}-B^{-1}\Phi'\Xi A\Phi=\mathrm{I}-B^{-1}M. \tag{17}$$

This shows that $\mathrm{I}-\Pi A$ is invertible if and only if the inverse $H^{-1}=(\mathrm{I}-B^{-1}M)^{-1}$ defining F exists.

When $(\mathrm{I}-\Pi A)^{-1}$ exists, based on (i) above, $(\mathrm{I}-\Pi A)^{-1}$ restricted to S coincides with the inverse of the mapping $(\mathrm{I}-\Pi A)|_S$, therefore $(\mathrm{I}-\Pi A)^{-1}$ maps S onto S, and furthermore, by Equation (16) and the existence of H^{-1},

$$(\mathrm{I}-\Pi A)^{-1}|_S=((\mathrm{I}-\Pi A)\big|_S)^{-1}=\Phi\cdot H^{-1}\cdot\Phi^{-1}. \tag{18}$$

Equation (15) then follows from $(\mathrm{I}-\Pi A)^{-1}=I+(\mathrm{I}-\Pi A)^{-1}\Pi A$ and Equations (18), (17) because

$$(I-\Pi A)^{-1}\Pi A=(\mathrm{I}-\Pi A)^{-1}\big|_S\cdot\Pi A=\Phi\cdot H^{-1}\cdot\Phi^{-1}\cdot\Pi A=\Phi F B^{-1}\Phi'\Xi A.$$

This completes the proof. □

Remark 2.1. The conclusion of the preceding lemma and its proof have analogous versions for general linear mappings of projection onto S, i.e., linear mappings Π with $\Pi^2=\Pi$ whose range is S. For example $\Pi=\Phi D$ where $D\Phi=\mathrm{I}$. This is because the proof does not rely on which type of projection Π is, so an analogous expression of $(\mathrm{I}-\Pi A)^{-1}$ for a general projection mapping can be derived similarly by substituting the matrix expression of Π in the proof. We also note that as mentioned earlier, B and M can be estimated by simulation and low-order calculations (see e.g., Bradtke and Barto [6], Boyan [5], Nedić and Bertsekas [13], Bertsekas and Yu [4]). Therefore, using the first part of Lemma 2.1, the existence of the inverse of $\mathrm{I}-\Pi A$ can also be verified using low-order calculations and simulation.

Lemma 2.2. *Let* H *and* D *be an* $n\times k$ *and* $k\times n$ *matrix, respectively. Let* $\|\cdot\|$ *denote the standard* (*unweighted*) *Euclidean norm. Then,*

$$\|HD\|_\xi^2=\|\Xi^{1/2}HD\Xi^{-1/2}\|^2=\sigma((H'\Xi H)(D\Xi^{-1}D')). \tag{19}$$

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 Equation (19) then follows from the definition of the norms: for any $n\times n$ matrix E,

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

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

For an $n \times n$ matrix E, we have $\|E\| = \sqrt{\sigma(E'E)}$. We proceed to prove the second equality in Equation (19) by studying the spectral radius of the symmetric positive semidefinite matrix $E'E$, where $E = \Xi^{1/2} H D \Xi^{-1/2}$. Define $W = H'\Xi H$ to simplify notation. We have

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

The second equality in Equation (19) is equivalent to the relation $\sigma(E'E) = \sigma(W(D\Xi^{-1}D'))$. The latter follows from Horn and Johnson [8, Theorem 1.3.20], but for completeness, we give a short direct proof. Let λ be a nonzero (necessarily real) eigenvalue of $E'E$, and let x be a nonzero corresponding eigenvector. We have

$$\Xi^{-1/2} D' W D \Xi^{-1/2} x = \lambda x, \tag{20}$$

so x is in $\text{col}(\Xi^{-1/2}D')$, the column space of $\Xi^{-1/2}D'$ and can be expressed as

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

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

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

Then, by Equation (20),

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

so that

$$\lambda r = W D \Xi^{-1} D' \bar{r} = W D \Xi^{-1} D' r. \tag{21}$$

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

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

proving the second equality in Equation (19). □

We now proceed to prove Theorem 2.1.

Proof of Theorem 2.1. To simplify notation, let us denote $y = x^* - \Pi x^*$ and $C = FB^{-1}$. By Equation (8) and the Pythagorean theorem,

$$\|x^* - \bar{x}\|_\xi^2 = \|y\|_\xi^2 + \|(\mathrm{I} - \Pi A)^{-1} \Pi A y\|_\xi^2. \tag{22}$$

It can be easily seen from the proof of Lemma 2.1 that $(\mathrm{I} - \Pi A)^{-1}\Pi = \Phi C \Phi' \Xi$. Applying Lemma 2.2 to the matrix $\Phi C \Phi' \Xi$ on the right-hand side with $H = \Phi C$ and $D = \Phi' \Xi$, we obtain

$$\|(\mathrm{I} - \Pi A)^{-1}\Pi\|_\xi^2 = \|HD\|_\xi^2 = \sigma(G), \quad \text{where } G = (H'\Xi H)(D\Xi^{-1}D').$$

Using this relation and the fact $\Pi = \Pi \cdot \Pi$, we can bound the second term on the right-hand side of Equation (22) by

$$\|(\mathrm{I} - \Pi A)^{-1}\Pi A y\|_\xi^2 = \|(\mathrm{I} - \Pi A)^{-1}\Pi \cdot \Pi A y\|_\xi^2 \le \sigma(G) \|\Pi A\|_\xi^2 \|y\|_\xi^2. \tag{23}$$

We also have

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

so G is the matrix G_1 given in the statement of the theorem. The bound (11) then follows by combining Equations (22) and (23). Finally, because $\sigma(G_1)$ is equal to $\|(\mathrm{I} - \Pi A)^{-1}\Pi\|_\xi^2$, the bound depends on the approximation subspace S and ξ and not the choice of Φ. □

We now prove Theorem 2.2.

Proof of Theorem 2.2. Let us denote $y = x^* - \Pi x^*$ and $C = FB^{-1}$. By Equation (8) and the Pythagorean theorem, we have

$$\|x^* - \bar{x}\|_\xi^2 = \|y\|_\xi^2 + \|(\mathrm{I} - \Pi A)^{-1}\Pi A y\|_\xi^2. \tag{24}$$

Similarly to the proof of Theorem 2.1, we proceed to bound the second term. Because

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

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

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

By Lemma 2.1,

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

Applying Lemma 2.2 to the matrix $\Phi C \Phi' \Xi A(\mathrm{I} - \Pi)$ on the right-hand side with $H = \Phi C$ and $D = \Phi' \Xi A(\mathrm{I} - \Pi)$, we obtain

$$\|(\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)\|_\xi^2 = \|HD\|_\xi^2 = \sigma(G), \quad \text{where } G = (H'\Xi H)(D\Xi^{-1}D'). \tag{26}$$

If we can establish that G is the matrix G_2 given in the statement of the theorem, then the bound (13) will follow by combining Equations (24), (25), and (26), and the invariance of the bound with respect to the choice of Φ will follow from Equation (26).

Indeed, we have

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

Using the definition $B = \Phi' \Xi \Phi$, we obtain $\Pi \Xi^{-1} = \Phi(\Phi' \Xi \Phi)^{-1}\Phi' \Xi \Xi^{-1} = \Phi B^{-1} \Phi'$, and it follows that

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

Combining the preceding two equations, we see that the matrix $D\Xi^{-1}D'$ is

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

with $R = \Phi' \Xi A \Xi^{-1} A' \Xi \Phi$. We also have

$$H' \Xi H = C' \Phi' \Xi \Phi C = C' B C,$$

so the matrix

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

is the matrix G_2 given in the statement of the theorem. This completes the proof. □

Remark 2.2. Lemmas 2.1 and 2.2 are useful for deriving other bounds besides Theorems 2.1 and 2.2. Several such bounds will be given in §3. In addition, we mention here that if the projection norm is $\|\cdot\|_\xi$ while the approximation error is measured with respect to a different norm $\|\cdot\|_{\tilde{\xi}}$, we can bound the error similarly, with the bounds expressed in terms of small-size matrices, as follows. Starting with the equality (8), we apply the triangle inequality instead of the Pythagorean theorem to obtain

$$\|x^* - \bar{x}\|_{\tilde{\xi}} \le \|x^* - \Pi x^*\|_{\tilde{\xi}} + \|(\mathrm{I} - \Pi A)^{-1}\Pi A(x^* - \Pi x^*)\|_{\tilde{\xi}},$$

and then bound the second term on the right-hand side using Lemmas 2.1 and 2.2. This argument can also be applied to the case where Π is a general linear mapping of projection onto S (cf. Remark 2.1).

2.2. Comparison of error bounds. The error bounds of Theorems 2.1 and 2.2 apply to the general case where ΠA is not necessarily a contraction mapping, while the contraction-based error bounds (3) and (4) apply only 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 noncontraction cases.

While the bounds (3), (4) can be derived in various equivalent ways, we will view them as relaxed versions of Equation (7),

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

In this way, we will place all the bounding approaches on an equal footing and be able to qualitatively compare them. In particular, we can obtain the bound (3), which is $\|x^* - \bar{x}\| \leq 1/(1-\alpha)\|x^* - \Pi x^*\|$ with $\alpha = \|\Pi A\|$, by writing

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

and by upper-bounding each term in the expansion separately: $\|(\Pi A)^n\| \leq \alpha^n$. We can obtain the bound (4), which is $\|x^* - \bar{x}\|_\xi \leq (1/\sqrt{1-\alpha^2})\|x^* - \Pi x^*\|_\xi$ with $\alpha = \|\Pi A\|_\xi$, by first writing

$$(\mathrm{I} - \Pi A)^{-1} = I + \Pi A(\mathrm{I} - \Pi A)^{-1}, \tag{27}$$

then by using the Pythagorean theorem and Equation (7) to obtain

$$\|x^* - \bar{x}\|_\xi^2 = \|x^* - \Pi x^*\|_\xi^2 + \|\Pi A(\mathrm{I} - \Pi A)^{-1}(x^* - \Pi x^*)\|_\xi^2 = \|x^* - \Pi x^*\|_\xi^2 + \|\Pi A(x^* - \bar{x})\|_\xi^2,$$

and finally by using the contraction property of ΠA to obtain

$$\|x^* - \bar{x}\|_\xi^2 \leq \|x^* - \Pi x^*\|_\xi^2 + \alpha^2\|x^* - \bar{x}\|_\xi^2$$

and rearranging terms. So equivalently, we can view the bound (4) as being derived by bounding the squared norm of the bias $\Pi A(\mathrm{I} - \Pi A)^{-1}(x^* - \Pi x^*)$ by $\alpha^2\|x^* - \bar{x}\|_\xi^2$, and relaxing the bound further to $(\alpha^2/(1-\alpha^2))\|x^* - \Pi x^*\|^2$, using the fact that ΠA is a contraction with modulus α. By contrast, the bounds of Theorems 2.1 and 2.2 are derived by bounding the norm of the equivalent expression $(\mathrm{I} - \Pi A)^{-1}\Pi A(x^* - \Pi x^*)$ of the bias directly, as shown earlier.

2.2.1. On the error bound of Theorem 2.2. We now show that the error bound of Theorem 2.2 is always sharper than the bound (4), and furthermore, for both contraction and noncontraction cases, the error bound of Theorem 2.2 is always the sharpest among all the bounds on $\|x^* - \bar{x}\|_\xi$ that do not depend on the vector b (Remark 2.3).

Proposition 2.1. *Let $\alpha = \|\Pi A\|_\xi$ and assume that $\alpha < 1$. Then, the error bound of Theorem 2.2 is always no worse than the error bound (4), i.e.,*

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

where G_2 is given by Equation (14).

Proof. Let $\gamma = \sqrt{\sigma(G_2)}$. If $\gamma = 0$, the statement is true. Consider now the case $\gamma > 0$. Because $\gamma = \sqrt{\sigma(G_2)} = \|(\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)\|_\xi$ by Theorem 2.2, we need to show that

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

Consider a vector $y \neq 0$ such that

$$\|(\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)y\|_\xi = \|(\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)\|_\xi \cdot \|y\|_\xi = \gamma\|y\|_\xi. \tag{28}$$

Then we must have $(\mathrm{I} - \Pi)y = y$, i.e., $\Pi y = 0$. (Otherwise, because $\gamma > 0$ and $y \neq 0$ implies $y \neq \Pi y$ by Equation (28), by redefining y to be $y - \Pi y \neq 0$, we can decrease $\|y\|_\xi$ while keeping the value of the left-hand side of (28) unchanged, which would contradict the definition of matrix norm.) Consider the following two equations in x,

$$x = (y - Ay) + Ax, \qquad x = \Pi(y - Ay) + \Pi Ax = \Pi Ax - \Pi Ay, \tag{29}$$

where the second equation is a projected equation corresponding to the first. Then, a solution of the first equation is $x^* = y$. Denote the solution of the second equation by $\bar{x}$. We have

$$\Pi y - \bar{x} = -\bar{x} = (\mathrm{I} - \Pi A)^{-1}\Pi A y = (\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)y, \tag{30}$$

and by Equation (28),

$$\|\Pi y - \bar{x}\|_\xi = \gamma\|y\|_\xi = \gamma\|y - \Pi y\|_\xi. \tag{31}$$

On the other hand, the error bound (4) for this case is equivalent to

$$\|\Pi y - \bar{x}\|_\xi^2 \le \left(\frac{1}{1-\alpha^2} - 1\right)\|y - \Pi y\|_\xi^2 = \frac{\alpha^2}{1-\alpha^2}\|y - \Pi y\|_\xi^2.$$

Together with Equation (31), this implies $\gamma^2 \le \alpha^2/(1-\alpha^2)$. □

Remark 2.3. The proof shows that regardless of whether ΠA is a contraction, the bound of Theorem 2.2 is tight, in the sense that for any A and S, there exists a worst-case choice of b for which the bound of Equation (13) in Theorem 2.2 holds with equality. This is the vector $b = y - Ay$ where y is any nonzero vector that satisfies Equation (28), as can be seen from the proof arguments in Equations (28)–(31).

2.2.2. On the error bound of Theorem 2.1. We now compare the error bound of Theorem 2.1 with the bounds (3) and (4). Because Theorem 2.1 is effectively equivalent to

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

we see that the bound of Theorem 2.1 is never worse than the bound (3), because we have bounded the norm of the matrix $(\mathrm{I} - \Pi A)^{-1}\Pi$ as a whole, instead of bounding each term in its expansion separately, as in the case of the bound (3).

However, the bound of Theorem 2.1 can be degraded by an over-relaxation: 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, unlike Theorem 2.2. This over-relaxation can have a significant effect when A has a dominant real eigenvalue $\beta \in (0, 1)$ with an eigenvector z that lies in the approximation subspace S. In such a case,

$$\|(\mathrm{I} - \Pi A)^{-1}\Pi z\|_\xi = \|(\mathrm{I} - \Pi A)^{-1} z\|_\xi = \frac{1}{1-\beta}\|z\|_\xi, \tag{32}$$

so we have

$$\|(\mathrm{I} - \Pi A)^{-1}\Pi\|_\xi \ge \frac{1}{1-\beta},$$

resulting in a large value of the bound if β is close to 1. By contrast, the residual vector $x^* - \Pi x^*$ cannot both be contained in S and be nonzero. Moreover, it can be shown that the bias in $\bar{x}$ is small if the residual vector is close to some eigenvector of A corresponding to a real eigenvalue, and in particular, there is no bias in $\bar{x}$ if the residual vector is such an eigenvector. [1] By contrast, as indicated by the discussion following Equation (32), the over-relaxation in the bounding procedure can have a pronounced effect when the approximation subspace S nearly contains the direction of an eigenvector of A associated with a real or complex eigenvalue close to 1, (where the relation of S with a complex eigenvector is considered in the complex vector space $\mathbb{C}^n$ with S being a subspace in the latter).

[1] When $x^* - \Pi x^*$ lies in the eigenspace of A corresponding to some real eigenvalue β, using the equation $\Pi A(x^* - \Pi x^*) = \beta\Pi(x^* - \Pi x^*) = 0$, we have

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

which shows that Πx^* satisfies the projected equation, so $\bar{x} = \Pi x^*$. Similarly, when $x^* - \Pi x^*$ is close to such an eigenspace of A, $\bar{x}$ and Πx^* are also close to each other. This can be seen as follows. Using $\Pi(x^* - \Pi x^*) = 0$, we have for any scalar β,

$$\Pi A(x^* - \Pi x^*) = \Pi(A - \beta I)(x^* - \Pi x^*).$$

Therefore, using Equation (9), the bias in $\bar{x}$ can be bounded by

$$\|\bar{x} - \Pi x^*\|_\xi \le \|(\mathrm{I} - \Pi A)^{-1}\Pi\|_\xi g(x^* - \Pi x^*), \quad \text{where } g(z) = \min_{\beta\in\Re}\|(A - \beta I)z\|_\xi.$$

The function g is continuous and vanishes at eigenvectors of A corresponding to real eigenvalues, which shows that if $x^* - \Pi x^*$ is close to such an eigenvector of A, then the bias in $\bar{x}$ is close to zero.

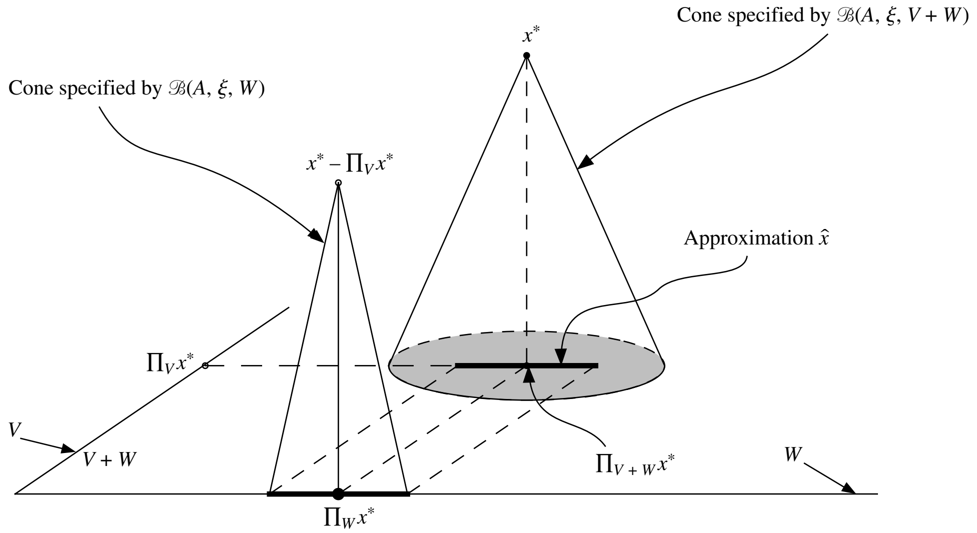


FIGURE 2. Illustration of Proposition 2.2 on transferring error bounds from one approximation subspace to another.
Notes. The subspaces V and W are such that $V \perp W$ and $\Pi_V x^*$ is known. The error bounds of Theorems 2.1 and 2.2 associated with the approximation subspace W can be transfered to $V+W$ 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 $\tilde{b} = b + A\Pi_V x^* - \Pi_V x^*$.

We thus consider ways of applying Theorem 2.1 that can be useful for obtaining sharper bounds under such circumstances. We describe one approach here, and a second one in the next section (Proposition 3.2). 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. This can be applied with Theorem 2.2 as well. We give a formal statement in the following proposition; see Figure 2 for an illustration.

In what follows, for a subspace V of $\Re^n$, let Π_V denote the projection on V with respect to $\|\cdot\|_\xi$. The inner product in $\Re^n$ is taken to be $\langle x, y\rangle_\xi = x'\Xi y$ for any $x, y \in \Re^n$, and orthogonality between vectors and subspaces of $\Re^n$ is accordingly defined with respect to $\langle\cdot,\cdot\rangle_\xi$.

PROPOSITION 2.2. *Let V and W be two orthogonal subspaces of $\Re^n$. Assume that $\Pi_V x^*$ is known and $\mathrm{I} - \Pi_W A$ is invertible. Let $\mathscr{B}(A, \xi, W)$ correspond to either the error bound of Theorem 2.1 or that of Theorem 2.2 with $S = W$. Then*

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

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

$$x = \Pi_W A x + \Pi_W \tilde{b}, \tag{33}$$

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

PROOF. First, notice that for any linear equation $x = Ax + b$, the error bounds of Theorems 2.1 and 2.2 do not depend on the vector b. Because $x^* - \Pi_V x^*$ 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

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

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

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

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

REMARK 2.4. We discuss implications of the proposition when V is an eigenspace of A, or more generally, an invariant subspace of A, namely, $A(V) \subset V$. In such a case, $A\Pi_V x^* \in V$, 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 (33) for $\bar{x}_w$. Then, we might 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, and correspondingly, $\Pi_V x^*$ is constant over all states and therefore an unimportant component of the cost vector x^*, which can be neglected in the process of policy iteration. When V is an eigenspace or invariant subspace of A, there is a simple relation between $\hat{x}$ and the solution $\bar{x}$ of the projected equation $x = \Pi_{V+W} b + \Pi_{V+W} Ax$. We have

$$\Pi_W \bar{x} = \Pi_W b + \Pi_W A(\Pi_W \bar{x} + \Pi_V \bar{x}) = \Pi_W b + \Pi_W A(\Pi_W \bar{x}),$$

so $\Pi_W \bar{x} = \Pi_W \hat{x} = \bar{x}_w$. Furthermore, because $\Pi_V \hat{x} = \Pi_V x^*$, the bias in $\hat{x}$ is alway no greater than the bias in $\bar{x}$ relative to $\Pi_{V+W} x^*$. (If V is not an invariant subspace of A, these no longer hold, which seems to imply that the bias in $\bar{x}$ might happen to be smaller than that in $\hat{x}$, although in computing $\bar{x}$ we do not use the knowledge of $\Pi_V x^*$.)

Remark 2.5. Proposition 2.2 also holds with $\Pi_V x^*$ replaced by any vector $v \in V$. In particular, we have

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

where $\hat{x} = v + \bar{x}_w$ and $\bar{x}_w$ is the solution of the projected equation $x = \Pi_W Ax + \Pi_W \tilde{b}$ 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 $\Pi_V x^*$.

We now mention another over-relaxation in the bound of Theorem 2.1. It has a less-pronounced effect than the first over-relaxation we just discussed but can degrade the bound in practice if ΠA is zero or near zero. When Theorem 2.1 is applied in practice, typically $\|\Pi A\|_\xi$ is unknown, and we substitute this term in the bound with some upper bound of it,[2] for instance, $\|A\|_\xi$ if the latter is known. Then, the bound we apply would be equivalent to

$$\|(\mathrm{I} - \Pi A)^{-1} \Pi A(x^* - \Pi x^*)\|_\xi \le \|(\mathrm{I} - \Pi A)^{-1} \Pi\|_\xi \|A\|_\xi \|x^* - \Pi x^*\|_\xi.$$

From the expression

$$(\mathrm{I} - \Pi A)^{-1} \Pi A = \Pi A + \Pi A (\mathrm{I} - \Pi A)^{-1} \Pi A, \tag{34}$$

we can see that when $\Pi A = 0$ and $A \ne 0$, the above matrix is zero but the bound

$$\|(\mathrm{I} - \Pi A)^{-1} \Pi\|_\xi \|A\|_\xi = \|\Pi + \Pi A(\mathrm{I} - \Pi A)^{-1} \Pi\|_\xi \|A\|_\xi = \sqrt{\sigma(G_1)} \|A\|_\xi$$

is not, as $\sigma(G_1) = \|\Pi\|_\xi^2 = 1$. Similarly, over-relaxation occurs when ΠA is not zero but is near zero, because the matrices Π and A are split in the bounding procedure.

The two shortcomings of the bound of Theorem 2.1 arise in the MDP applications that we will discuss, as well as in more general contexts with noncontraction mappings. On the other hand, even for contraction mappings, there are cases where Theorem 2.1 provides sharper bounds than the fixed error bound (4), and cases where Theorem 2.1 gives computable bounds while the bound (4) provides no quantitative information—for example, when ΠA is a contraction but with the modulus of contraction being unknown, as is the case with policy evaluation in MDP under the average cost criterion (see §4).

3. Related results.

In this section, we first give two error bounds (Propositions 3.1 and 3.2) for projected equations that are in decomposition forms. The first bound is expressed in terms of components of the residual vector $x^* - \Pi x^*$ in mutually orthogonal subspaces, and the second bound is a bound on a component of the approximation error $x^* - \bar{x}$ in some subspace of interest. We then derive an analogous computable error bound (Proposition 3.3) for an alternative approximation approach, namely, the equation error minimization method. Our line of analysis is similar to the one in §2. In particular, Propositions 3.1, 3.2, and 3.3 require applications of Lemma 2.2 with different matrices, just like Theorems 2.1 and 2.2.

3.1. Error bounds in decomposition forms for projected equations.

We first investigate how the component of the residual vector $x^* - \Pi x^*$ in some subspace could 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 projected equations corresponding to TD(λ) with $\lambda > 0$), as in Theorem 2.1, while the analysis takes into account that $x^* - \Pi x^*$ is orthogonal to S, like in Theorem 2.2. Because of the latter, the bound is not susceptible to the over-relaxation issues in the bound of Theorem 2.1 as discussed in §2.2.2.

[2] We note a subtlety here relating to the computation of the bound of Theorem 2.1 in practice. As an application of Lemma 2.2, the term $\|\Pi A\|_\xi$ can be expressed in terms of the spectral radius of a small-size matrix, which can be estimated using simulation for certain cases. However, such calculation involves the same procedure as estimating the matrix R in the calculation of the bound of Theorem 2.2, while Theorem 2.2 supersedes Theorem 2.1. Hence, if we can carry out such calculation, we should use Theorem 2.2 instead of Theorem 2.1.

Proposition 3.1. *Let $\hat{S}_i$, $i = 1, \ldots, m$, be m mutually orthogonal subspaces of $\Re^n$ such that*

$$x^* - \Pi x^* \in \hat{S}_1 + \hat{S}_2 + \cdots + \hat{S}_m.$$

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

$$\|\bar{x} - \Pi x^*\|_\xi \le \sum_{i=1}^m \sqrt{\sigma(\hat{G}_i)}\,\|\hat{\Pi}_i(x^* - \Pi x^*)\|_\xi, \tag{35}$$

where $\hat{\Pi}_i$ is the mapping of projection on $\hat{S}_i$ with respect to $\|\cdot\|_\xi$, and $\hat{G}_i$ is the $\hat{k}_i \times \hat{k}_i$ matrix

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

with

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

and

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

Furthermore, $\sigma(\hat{G}_i) = \|(\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)\hat{\Pi}_i\|_\xi^2$, so the bound (35) is invariant with respect to the choice of basis vectors of $\hat{S}_i$, $i = 1, \ldots, m$, and S (i.e., Ψ_i, $i = 1, \ldots, m$, and Φ).

Proof. Our line of analysis is the same as the one for Theorem 2.2. Let us denote $y = x^* - \Pi x^*$ and $C = FB^{-1}$. The assumption implies that $y = \sum_{i=1}^m \hat{\Pi}_i y$. Together with Equation (9), this shows

$$\Pi x^* - \bar{x} = (\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)y = \sum_{i=1}^m (\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)\hat{\Pi}_i y. \tag{36}$$

By Lemma 2.1 and the definition of $\hat{\Pi}_i$,

$$(\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)\hat{\Pi}_i = \Phi C \Phi'\Xi A(\mathrm{I} - \Pi)\Psi_i \hat{B}_i^{-1}\Psi_i'\Xi. \tag{37}$$

Applying Lemma 2.2 to the matrix on the right-hand side with

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

we have

$$\|(\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)\hat{\Pi}_i\|_\xi = \sqrt{\sigma(G)}, \tag{38}$$

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

$$\|(\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)\hat{\Pi}_i y\|_\xi = \|(\mathrm{I} - \Pi A)^{-1}\Pi A(\mathrm{I} - \Pi)\hat{\Pi}_i \cdot \hat{\Pi}_i y\|_\xi \le \sqrt{\sigma(G)}\,\|\hat{\Pi}_i y\|_\xi. \tag{39}$$

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

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

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

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

Combining Equations (36), (39), and the triangle inequality, we obtain the bound (35) and its invariance with respect to the choice of basis vectors of S and $\hat{S}_i$, $i = 1, \ldots, m$. □

In the above bound, for all i, the matrices $\hat{B}_i$ and $E_{i,2}$ have similar forms to the matrix B, and the matrices $E_{i,1}$ have similar forms to the matrix M. The procedure for estimating B and M can be used directly to estimate these matrices. In particular, to estimate $E_{i,1}$ or $E_{i,2}$ we can apply the estimation procedure for $M = \Phi'\Xi A \Phi$ or $B = \Phi'\Xi\Phi$, respectively, with the second matrix Φ replaced by Ψ_i; and similarly, to estimate $\hat{B}_i$, we can apply the procedure for B with Φ replaced by Ψ_i.

Although it is generally impractical to estimate all terms involved in the above bound, it might be of interest in special cases to calculate some of the individual terms. For example, suppose the subspace S is close to x^*, so the residual $x^* - \Pi x^*$ is small. Then, if some term $\sigma(\hat{G}_i)$ is significantly larger than 1, we might consider including the corresponding subspace $\widehat{S}_i$ in the approximation subspace to prevent a potentially large bias in the approximating solution caused by the residual component $\widehat{\Pi}_i(x^* - \Pi x^*)$ in the worst case.

A special choice of S often used in practice is the one resulting from a hard aggregation scheme: we partition $\{1, 2, \ldots, n\}$ into k subsets and let S be the set of vectors that are constant over each subset. Then, $\widehat{S}_i$ can naturally be chosen to correspond to a finer partition of some subset.

We now bound a component of the approximation error, $\Pi_{\hat{S}}(x^* - \bar{x})$, in some subspace $\hat{S}$, where $\Pi_{\hat{S}}$ denotes projection on $\hat{S}$ with respect to $\|\cdot\|_\xi$. This type of bound can be useful when we are interested in the approximation error in a small subset of entries of $\bar{x}$; in the MDP context, these entries correspond to a subset of states.

We observe from Equation (8) that

$$\Pi_{\hat{S}}(x^* - \bar{x}) = \Pi_{\hat{S}}(x^* - \Pi_S x^*) + \Pi_{\hat{S}}(\mathrm{I} - \Pi_S A)^{-1}\Pi_S A(x^* - \Pi_S x^*). \tag{40}$$

Similarly to the derivation of Theorem 2.1, we can bound the first term by

$$\|\Pi_{\hat{S}}(x^* - \Pi_S x^*)\|_\xi = \|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S)(x^* - \Pi_S x^*)\|_\xi \le \|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S)\|_\xi \|x^* - \Pi_S x^*\|_\xi,$$

and bound the second term by

$$\|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S A)^{-1}\Pi_S A(x^* - \Pi_S x^*)\|_\xi \le \|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S A)^{-1}\Pi_S\|_\xi \|\Pi_S A\|_\xi \|x^* - \Pi_S x^*\|_\xi.$$

We then apply Lemmas 2.1 and 2.2 to express the matrix norms $\|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S)\|_\xi$, $\|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S A)^{-1}\Pi_S\|_\xi$ in the above bounds as the spectral radii of small-size matrices. Applying the triangle inequality with Equation (40) then gives us a bound on $\|\Pi_{\hat{S}}(x^* - \bar{x})\|_\xi$.

Alternatively, we can derive another bound on $\|\Pi_{\hat{S}}(x^* - \bar{x})\|_\xi$ similarly to the derivation of Theorem 2.2 as follows. We start with the equality relation

$$\Pi_{\hat{S}}(x^* - \bar{x}) = \Pi_{\hat{S}}(\mathrm{I} - \Pi_S A)^{-1}(x^* - \Pi_S x^*),$$

which follows from Equation (7) and is equivalent to Equation (40). We then bound $\|\Pi_{\hat{S}}(x^* - \bar{x})\|_\xi$ using the fact that

$$\begin{aligned}\|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S A)^{-1}(x^* - \Pi_S x^*)\|_\xi &= \|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S A)^{-1}(\mathrm{I} - \Pi_S)(x^* - \Pi_S x^*)\|_\xi \\ &\le \|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S A)^{-1}(\mathrm{I} - \Pi_S)\|_\xi \|x^* - \Pi_S x^*\|_\xi,\end{aligned}$$

and we apply Lemma 2.1 to obtain the expression $(\mathrm{I} - \Pi_S A)^{-1}(\mathrm{I} - \Pi_S) = I + \Phi F B^{-1}\Phi'\Xi(A - I)$, and we subsequently apply Lemma 2.2 to express the matrix norm $\|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S A)^{-1}(\mathrm{I} - \Pi_S)\|_\xi$ as the spectral radius of a small-size matrix.

We state in the following proposition the two bounds on $\|\Pi_{\hat{S}}(x^* - \bar{x})\|_\xi$ that we just discussed, omitting the algebraic details of the proof.

Proposition 3.2. *Let $\hat{S}$ be a subspace of $\Re^n$ and Ψ be an $n \times \hat{k}$ matrix whose columns form a basis of $\hat{S}$. Let*

$$\widehat{B} = \Psi'\Xi\Psi, \qquad \widehat{E} = \Phi'\Xi\Psi, \qquad \widehat{M} = \Phi'\Xi A\Psi, \qquad \widehat{C} = \widehat{E}'FB^{-1}.$$

Then

$$\|\Pi_{\hat{S}}(x^* - \bar{x})\|_\xi \le \left(\sqrt{\sigma(\bar{G}_1)} + \sqrt{\sigma(\bar{G}_2)}\|\Pi_S A\|_\xi\right)\|x^* - \Pi_S x^*\|_\xi, \tag{41}$$

where $\bar{G}_1$ and $\bar{G}_2$ are $\hat{k} \times \hat{k}$ matrices given by

$$\bar{G}_1 = \mathrm{I} - \widehat{B}^{-1}\widehat{E}'B^{-1}\widehat{E}, \quad \bar{G}_2 = \widehat{B}^{-1}\widehat{C}F'\widehat{E}.$$

Also,

$$\|\Pi_{\hat{S}}(x^* - \bar{x})\|_\xi \le \sqrt{\sigma(\bar{G}_3)}\|x^* - \Pi_S x^*\|_\xi, \tag{42}$$

where $\bar{G}_3$ is a $\hat{k} \times \hat{k}$ matrix given by

$$\bar{G}_3 = I + \bar{G}_2 + \widehat{B}^{-1}(W + W') + \widehat{B}^{-1}\widehat{C}(R - M - M')\widehat{C}',$$

with

$$W = \hat{C}(\hat{M} - \hat{E}),$$

and R as given in Theorem 2.2. Furthermore,

$$\sigma(\bar{G}_1) = \|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S)\|_\xi^2, \quad \sigma(\bar{G}_2) = \|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S A)^{-1}\Pi_S\|_\xi^2, \quad \sigma(\bar{G}_3) = \|\Pi_{\hat{S}}(\mathrm{I} - \Pi_S A)^{-1}(\mathrm{I} - \Pi_S)\|_\xi^2,$$

so the bounds (41), (42) *are invariant with respect to the choice of basis vectors of $\hat{S}$ and S (i.e., Ψ and Φ).*

In the preceding proposition, the matrices $\hat{B}$ and $\hat{E}$ have similar forms to the matrix B, while the matrix $\hat{M}$ has a similar form to the matrix M. Therefore, they can be estimated using respective procedures for estimating B and M with simulation, as explained earlier after the proof of Proposition 3.1. Estimating the bound (41) is thus straightforward for all TD-type methods, while estimating the bound (42) is less so because it involves the matrix R given in Theorem 2.2. When the approximation subspace S nearly contains the direction of some eigenvector of A associated with an eigenvalue close to 1, but the subspace $\hat{S}$ of interest does not, the bound (41) may be used as an alternative way besides Proposition 2.2 to bypass the eigenspace-related over-relaxation issue in the bound of Theorem 2.1 as discussed in §2.2.2.

3.2. Error bound for an alternative approximation method. Our line of analysis in §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

$$\min_{x \in S} \|x - Ax - b\|_\xi^2 \tag{43}$$

as an approximation of x^*. We will make the assumption that $\mathrm{I} - A$ is invertible, under which both x^* and the solution of (43) are unique.

Proposition 3.3. *Assume $\mathrm{I} - A$ is invertible. Let $\tilde{x}$ be the solution of the minimization problem* (43). *Then*

$$\|x^* - \tilde{x}\|_\xi \le \sqrt{1 + \sigma(\tilde{G})}\,\|x^* - \Pi x^*\|_\xi, \tag{44}$$

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

$$\tilde{G} = B\tilde{E}\tilde{R}\tilde{E} - I, \tag{45}$$

with

$$B = \Phi'\Xi\Phi, \qquad \tilde{E} = (\Phi' L'\Xi L\Phi)^{-1}, \qquad \tilde{R} = \Phi' L'\Xi L \Xi^{-1} L'\Xi L\Phi, \tag{46}$$

and $L = \mathrm{I} - A$. Furthermore, the bound (44) *is invariant with respect to the choice of basis vectors of S (i.e., Φ).*

Proof. The bound is equivalent to the bound on the bias: $\|\tilde{x} - \Pi x^*\|_\xi \le \sqrt{\sigma(\tilde{G})}\,\|x^* - \Pi x^*\|_\xi$. First, we establish an equality relation analogous to Equations (7) and (9):

$$\tilde{x} - \Pi x^* = C(x^* - \Pi x^*), \quad \text{where } C = \Phi(\Phi' L'\Xi L\Phi)^{-1}\Phi' L'\Xi L. \tag{47}$$

We then apply Lemma 2.2, taking into account that $\Pi(x^* - \Pi x^*) = 0$, similarly to the analysis for Theorem 2.2.

The minimization problem (43) can be written as $\min_{r \in \Re^k} \|L\Phi r - b\|_\xi^2$. The optimality condition is

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

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

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

Subtracting the last two equations, we have

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

Because 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 Equation (47).

Because $\Pi(x^* - \Pi x^*) = 0$, Equation (47) is further equivalent to

$$\tilde{x} - \Pi x^* = \Phi(\Phi' L' \Xi L \Phi)^{-1} \Phi' L' \Xi L (\mathrm{I} - \Pi)(x^* - \Pi x^*). \tag{48}$$

Therefore,

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

Applying Lemma 2.2 to the matrix in the right-hand side with

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

we have

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

where $G = (H'\Xi H)(D\Xi^{-1}D') = B(D\Xi^{-1}D')$. Similarly to the calculation in the proof of Theorem 2.2, it can be shown that

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

where

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

Thus,

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

which is the matrix $\tilde{G}$ given in Equation (45).

Finally, we prove that the bound is invariant with respect to the choice of basis vectors of S. To see this, we write the matrix $\Phi(\Phi' L' \Xi L \Phi)^{-1} \Phi' L' \Xi L (\mathrm{I} - \Pi)$ in Equation (48) equivalently as the product of three matrices: $L^{-1} \cdot L\Phi(\Phi' L' \Xi L \Phi)^{-1} \Phi' L' \Xi \cdot L(\mathrm{I} - \Pi)$. 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 $L(S)$ with respect to $\|\cdot\|_\xi$, hence it also does not depend on the choice of Φ. Therefore, the bound is invariant with respect to the choice of Φ. □

Similarly to the argument in the proof of Proposition 2.1 [cf. Remark 2.3], one can show that the bound of Proposition 3.3 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. One could also use an argument similar to that for Proposition 3.2 to bound a component of the approximation error, $\Pi_{\hat{S}}(x^* - \tilde{x})$, in some subspace $\hat{S}$ of interest.

We also note an alternative expression of the bound of Proposition 3.3 and an alternative way of proof. We have $1 + \sigma(\tilde{G}) = \sigma(\tilde{G} + I)$ because the eigenvalues of the matrix $\tilde{G}$ are all real and nonnegative, as shown in the proof of Lemma 2.2. We thus obtain $\sqrt{1 + \sigma(\tilde{G})} = \sqrt{\sigma(B\tilde{E}\tilde{R}\tilde{E})}$. As B. Scherrer pointed out to us, one may prove the bound alternatively as follows. We start with the relation $\tilde{x} - x^* = (C - I)(x^* - \Pi x^*)$, where the matrix C is given in Equation (47), exploit the idempotent property of C (i.e., $C^2 = C$) and its implication that $\|C - I\|_\xi = \|C\|_\xi$ when C is neither the identity nor the zero matrix (see e.g., Szyld [16]), and then proceed with Lemma 2.2 to obtain the bound.

The matrices $\tilde{E}^{-1}$ and $\tilde{R}$ have similar forms to the matrices B and R, respectively, with the matrix $L\Phi$ in place of the matrix Φ in B and R. This shows that the procedures for estimating B and R can be applied with slight modifications to estimate $\tilde{E}^{-1}$ and $\tilde{R}$.

4. Applications.

We consider two applications of Theorems 2.1 and 2.2. 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 optimal cost approximation for optimal stopping problems. The second application is approximating solutions of large general systems of linear equations. We also illustrate with figures various issues discussed in §2.2 on the comparison of the bounds.

4.1. Cost function approximation for MDP.

We start with the case of policy evaluation in MDP and consider the case of optimal cost approximation in stopping problems in §4.1.3. For policy evaluation, 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. For simplicity of discussion, we assume that the Markov chain is irreducible. The original linear equation that we want to solve is the Bellman equation, or optimality equation, satisfied by x^*. It takes the form

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

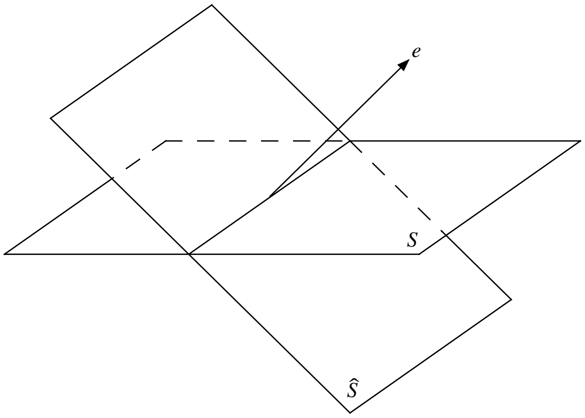


FIGURE 3. Illustration of $\hat{S}$, the orthogonal complement of V_e in $S+V_e$ with $V_e = \text{Span}\{e\}$, i.e., $\hat{S} = (S+V_e)\cap V_e^{\perp}$.

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).

With the TD(λ) method, we solve a projected form of the multistep Bellman equation parametrized by $\lambda \in [0,1]$ and satisfied by x^*. Let T be the Bellman operator given by $T(x) = g + \alpha P x$. The multistep Bellman equation corresponding to λ and its projected form solved by TD(λ) are given by

$$x = T^{(\lambda)}(x) = b + Ax, \qquad x = \Pi b + \Pi A x, \tag{49}$$

respectively, where the mapping $T^{(\lambda)}$, the matrix A, and the vector b are defined for a pair of values (α,λ) with $\lambda\in[0,1]$ for $\alpha\in[0,1)$ and $\lambda\in[0,1)$ for $\alpha=1$ by

$$T^{(\lambda)}(x) = (1-\lambda)\sum_{m=0}^{\infty}\lambda^m T^{m+1}(x), \quad \lambda\in[0,1), \qquad T^{(1)}(x) = \lim_{\lambda\to 1} T^{(\lambda)}(x),$$

and

$$A = P^{(\alpha,\lambda)} \overset{\text{def}}{=} (1-\lambda)\sum_{m=0}^{\infty}\lambda^m(\alpha P)^{m+1}, \qquad b = \sum_{m=0}^{\infty}\lambda^m(\alpha P)^m g. \tag{50}$$

(We omit the case $\alpha = 1, \lambda = 1$ for simplicity of discussion; it requires additional notation but can be dealt with similarly.) Notice that when $\lambda = 0$, we have $A = \alpha P$ and $b = g$, and TD(0) solves the projected Bellman equation. On the other hand, when $\lambda = 1$ and $\alpha < 1$, A is the zero matrix and $b = x^*$, (so the projected equation is vacuous), and TD(1) computes the projection of the solution x^*. In general, when λ is between zero and one, there is a trade-off between the potential bias in the approximating solution $\bar{x}$ and the variance of the estimators of $\bar{x}$ using simulated trajectories of the Markov chain. The increase in the variance as λ approaches 1 is in part due to the increasing variance of estimators of terms involving $\lambda^m\alpha^m P^m$ for large values of m, which appear in the definitions of A and b. This trade-off is known in practice and widely discussed in the literature (see e.g., Konda [9] and references therein).

We note that for the multistep projected Bellman equation solved by TD(λ) with $\lambda > 0$, we do not yet have an efficient simulation-based method for estimating the bound of Theorem 2.2; we have calculated the bound using common matrix algebra, and we plot it just for comparison.

4.1.1. Discounted problems. Consider the discounted case: $\alpha < 1$. First, we compare the bounds for the case where ξ is the invariant distribution of the Markov chain. For $\lambda \in [0,1]$, the modulus of contraction of $A = P^{(\alpha,\lambda)}$ with respect to $\|\cdot\|_\xi$ is

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

For any approximation subspace S, we can upper bound $\|\Pi P^{(\alpha,\lambda)}\|_\xi$ by $(1-\lambda)\alpha/(1-\lambda\alpha)$ in the standard bounds (3), (4) and the bound of Theorem 2.1 to obtain the bounds as a function of λ. This extra upper bounding step is not needed in applying Theorem 2.2. Figure 4 illustrates the error bounds. It can be observed from Figure 4(b) and 4(d) that the bound of Theorem 2.2 has consistently performed best, as indicated by the analysis.

Let e denote the constant vector of all ones. The stochastic matrix P has a dominant eigenvalue 1 with e being an associated eigenvector. Like P, the matrix $A = P^{(\alpha,\lambda)}$ has e as an eigenvector associated with the dominant eigenvalue $(1-\lambda)\alpha/(1-\lambda\alpha)$. As discussed in §2.2.2, if the approximation subspace S contains or nearly contains the direction of the eigenvector e of $P^{(\alpha,\lambda)}$, the bound of Theorem 2.1 can degrade. In the case here, it can degrade essentially to the standard error bound (3), because $\|P^{(\alpha,\lambda)}\|_\xi$ coincides with the spectral radius of $P^{(\alpha,\lambda)}$. To get around this issue, we can consider an alternative way of estimating x^* and obtaining sharper bounds, which is an application of Proposition 2.2 and Remark 2.4 in §2.2.2. We give the details in what follows and discuss in what sense the bounds are comparable to each other.

In particular, let $V_e = \text{Span}\{e\}$. We can estimate separately the projection of x^* on V_e and the projection of x^* on another subspace $\hat{S} = (S+V_e)\cap V_e^\perp$, which is the orthogonal complement of V_e in $S+V_e$ (see Figure 3), and define the sum of the two estimates to be the approximating solution $\hat{x}$. More precisely, this is an application of Proposition 2.2 and Remark 2.4 with $V=V_e$ and $W=\hat{S}$. Let Π_V for a subspace V denote the projection on V with respect to $\|\cdot\|_\xi$. We define $\hat{x} = \Pi_{V_e}x^* + \bar{x}_{\hat{S}}$, where $\bar{x}_{\hat{S}}$ satisfies the projected equation

$$x = \Pi_{\hat{S}}(b + A\Pi_{V_e}x^* - \Pi_{V_e}x^*) + \Pi_{\hat{S}}Ax = \Pi_{\hat{S}}b + \Pi_{\hat{S}}Ax, \tag{51}$$

which is a projected version of the equation $x = (b + A\Pi_{V_e}x^* - \Pi_{V_e}x^*) + Ax$ satisfied by $x^* - \Pi_{V_e}x^*$. This approach can be easily implemented with simulation in practice.[3] We have that the error bound of Theorem 2.1 or 2.2 for the projected Equation (51) associated with $\hat{S}$ carries over to the combined estimate $\hat{x} = \Pi_{V_e}x^* + \bar{x}_{\hat{S}} \in S+V_e$:

$$\|\hat{x}-x^*\|_\xi \le \mathscr{B}(A,\xi,\hat{S})\|x^* - \Pi_{S+V_e}x^*\|_\xi \le \mathscr{B}(A,\xi,\hat{S})\|x^*-\Pi_S x^*\|_\xi,$$

where the second inequality follows from $\|x^* - \Pi_{S+V_e}x^*\|_\xi \le \|x^*-\Pi_S x^*\|_\xi$. Comparing this bound with an error bound on the solution $\bar{x}$ of the projected equation associated with S,

$$\|\bar{x}-x^*\|_\xi \le \mathscr{B}(A,\xi,S)\|x^*-\Pi_S x^*\|_\xi,$$

we see that it is sensible to compare the bound $\mathscr{B}(A,\xi,\hat{S})$ with $\mathscr{B}(A,\xi,S)$: we interpret the former (latter) as that one can construct an approximating solution in $S+V_e$ (S) with the approximation error being no greater than $\mathscr{B}(A,\xi,\hat{S})\|x^*-\Pi_S x^*\|$ $(\mathscr{B}(A,\xi,S)\|x^*-\Pi_S x^*\|_\xi)$. If $\mathscr{B}(A,\xi,\hat{S}) \le \mathscr{B}(A,\xi,S)$, it would imply that in terms of approximation error, $\hat{x}$ has a better worst-case guarantee than $\bar{x}$. (If $e\in S$, then it always holds that $\hat{x}$ has less bias than $\bar{x}$; see Remark 2.4.)

Figure 4 shows how the use of $\hat{S}$ may improve the error bounds, and the improvement is significant in this case for applying Theorem 2.1. Figure 4(d) illustrates that the bound of Theorem 2.1 for the projected equation associated with $\hat{S}$ can still be too loose, if S nearly contains the direction of some other eigenvector of A besides e that is associated with an eigenvalue close to 1. (Here, the eigenvector and eigenvalue can be complex, and the relation of S with the eigenvector is considered in $\mathbb{C}^n$ with S being a subspace in the latter.) Proposition 2.2 and Remarks 2.4, 2.5 are, in principle, applicable in such cases, but there is a difficulty in practice to obtain the eigenvectors of A other than the trivial one, e. Figure 4(d) also illustrates that Theorem 2.2 is significantly less affected by this eigenspace issue.

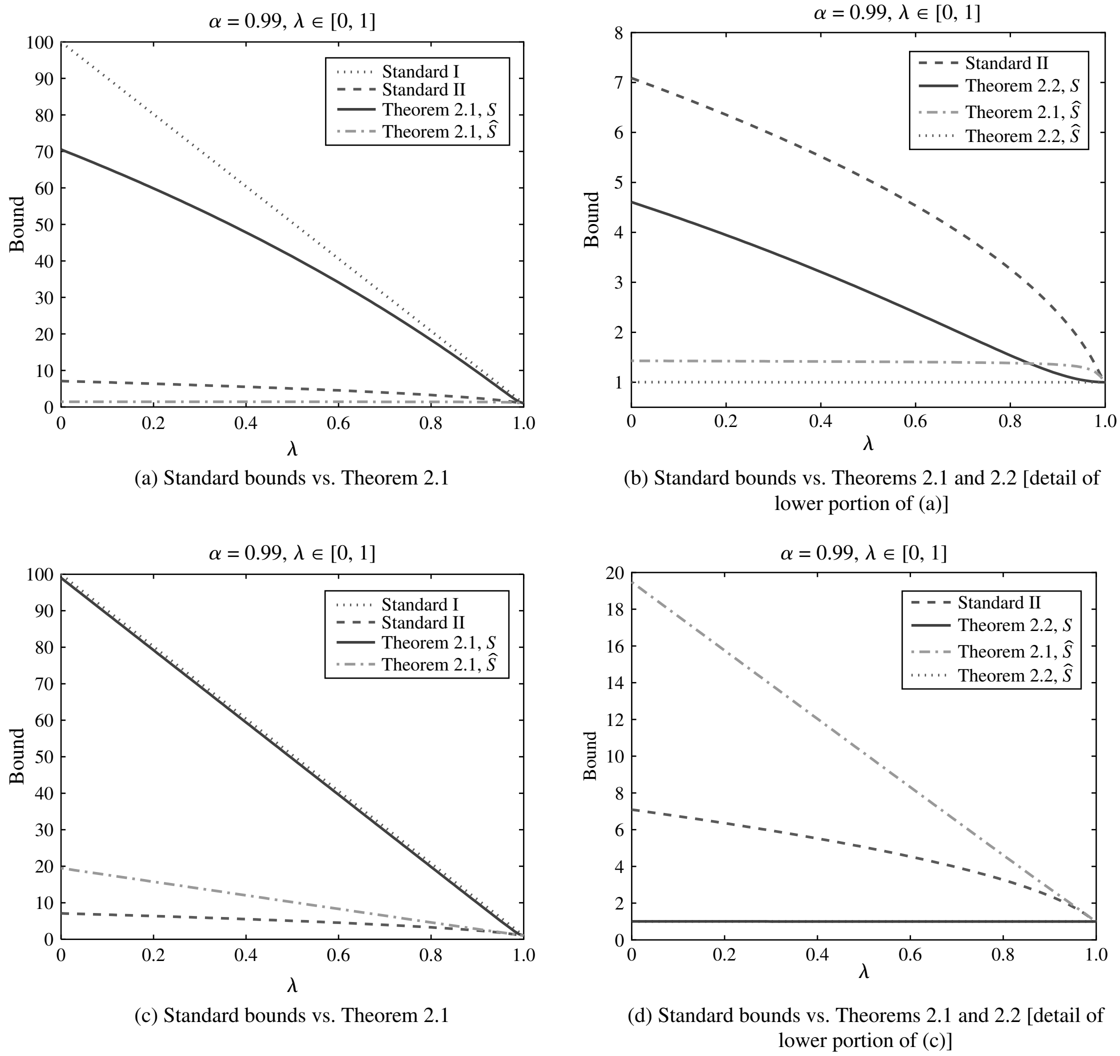
Figure 5 compares the bounds for the case where the projection norm is the standard unweighted Euclidean norm. With respect to this norm, the mapping $\Pi A = \Pi P^{(\alpha,\lambda)}$ is not necessarily a contraction for small values of λ, even though in the example in Figure 5 it is. The standard bounds and the bound of Theorem 2.1 need the value $\|\Pi A\|$, while the bound of Theorem 2.2 does not. For comparison of these bounds, we compute $\|P\|$ using the knowledge of P, bound $\|\Pi A\|$ by $(1-\lambda)\|\alpha P\|/(1-\lambda\|\alpha P\|)$, and plug the latter in the standard bounds and the bound of Theorem 2.1. The value $\|\alpha P\|$ is shown in the titles of the subfigures in Figure 5. The behavior of the bounds is similar to that in Figure 4.

Note that although an upper bound on $\|\Pi A\|_\xi$ is needed to apply Theorem 2.1, similar to when applying the standard error bounds, Theorem 2.1 is applicable in cases where the best upper bound on $\|\Pi A\|_\xi$ we know is no less than 1, whereas the standard error bounds are inapplicable in such cases.

[3] Implementing this approach in practice is straightforward, and we note some details here. The basis vectors of $\hat{S}$ can be generated from Φ by subtracting $\xi'\Phi$ from the rows of Φ. These row vectors are referred to as feature vectors associated with the states. The vector $\xi'\Phi$ is simply the mean feature vector when the random process of states is stationary, with the stationary distribution given by ξ, so $\xi'\Phi$ can be estimated easily by sample average. This procedure of changing the approximation subspace by subtracting the mean feature is common in the context of average cost MDP problems; see, e.g., Konda [9]. We can also compute easily the projection of x^* on V_e when ξ is the invariant distribution of the Markov chain associated with P; in particular,

$$\xi'x^* = \xi'g + \alpha\xi'Px^* = \xi'g + \alpha\xi'x^* \quad\Rightarrow\quad p\Pi_{V_e}x^* = \frac{\xi'g}{1-\alpha}e,$$

so $\Pi_{V_e}x^*$ can be calculated through simulation by averaging the one-stage costs. When ξ is not the invariant distribution associated with P, it is difficult to compute $\Pi_{V_e}x^*$ without bias. However, this component is in any case not absolutely needed, because it is constant over all states and not useful for policy iteration, and also because it is not needed in the projected Equation (51), as noted in Remark 2.4 in §2.2.2.



(a) Standard bounds vs. Theorem 2.1

(b) Standard bounds vs. Theorems 2.1 and 2.2 [detail of lower portion of (a)]

(c) Standard bounds vs. Theorem 2.1

(d) Standard bounds vs. Theorems 2.1 and 2.2 [detail of lower portion of (c)]

FIGURE 4. Comparison of error bounds as functions of λ for two discounted problems with randomly generated Markov chains.

Notes. The dimension parameters are $n = 200$, $k = 50$, and the weights ξ in the projection norm is the invariant distribution. Standard I and Standard II refer to the bounds (3) and (4), 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, causing P to have a subdominant eigenvalue relatively close to 1; S is chosen to contain e and a vector close to an eigenvector associated with that subdominant eigenvalue. The subspace $\hat{S}$ is derived from S by orthogonalization, as shown in Figure 3. In (d), the bounds of Theorem 2.2 corresponding to the two subspaces S and $\hat{S}$ coincide.

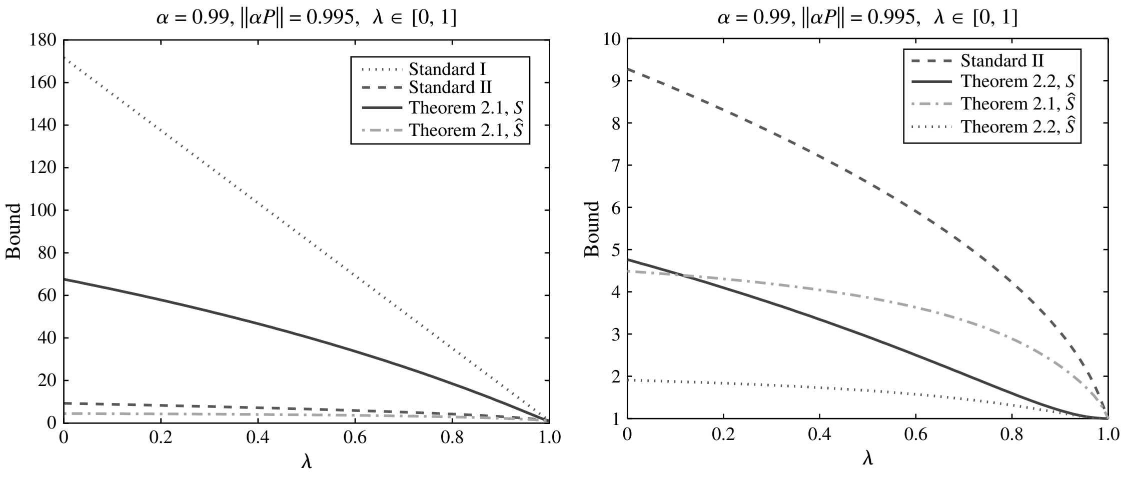
As a particular application example related to this discussion, let us consider the following scheme, which is commonly used for exploring states and actions that do not occur frequently under the policy to be evaluated. At each time, at some state, we follow that policy with probability $\rho < 1$ and deviate from it with probability $1-\rho$ for the purpose of exploration. This induces a Markov chain whose transition matrix is of the form $\rho P + (1-\rho)\tilde{P}$ for some stochastic matrix $\tilde{P} \neq P$. Correspondingly, we let the weights ξ in the projection norm be the invariant distribution of this Markov chain, i.e.,

$$\xi'(\rho P + (1-\rho)\tilde{P}) = \xi'. \tag{52}$$

Then, for the projected equation solved by TD(λ) and $\rho > \lambda^2\alpha^2$, it is not difficult to show by using the implication of the above relation, $\xi' P \leq \xi'/\rho$, that we can simply bound the matrix norms $\|P\|_\xi$ and $\|\Pi A\|_\xi = \|\Pi P^{(\alpha,\lambda)}\|_\xi$ by

$$\|P\|_\xi \leq \frac{1}{\sqrt{\rho}}, \qquad \|\Pi P^{(\alpha,\lambda)}\|_\xi \leq \|P^{(\alpha,\lambda)}\|_\xi \leq \frac{(1-\lambda)\|\alpha P\|_\xi}{1-\lambda\|\alpha P\|_\xi} \leq \frac{(1-\lambda)\alpha/\sqrt{\rho}}{1-\lambda\alpha/\sqrt{\rho}}, \tag{53}$$

where neither $\|\Pi P^{(\alpha,\lambda)}\|_\xi$ nor its upper bound above is necessarily less than 1. (For example, with $\lambda = 0$ and $\rho = 1/2$, we have by Equation (53) $\|\Pi P^{(\alpha,\lambda)}\|_\xi \leq \|\alpha P\|_\xi \leq \sqrt{2}\alpha$.) With the upper bound on $\|\Pi P^{(\alpha,\lambda)}\|_\xi$ given



(a) Standard bounds vs. Theorem 2.1

(b) Standard bounds vs. Theorems 2.1 and 2.2 [detail of lower portion of (a)]

FIGURE 5. Comparison of error bounds for discounted problems.

Notes. The setup is the same as 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.

in Equation (53), we can apply Theorem 2.1 to compute an error bound on $\|x^* - \bar{x}\|_\xi$ for the above exploration scheme with $\rho > \lambda^2\alpha^2$.

Thus the availability of computable error bounds for noncontraction mappings can facilitate the design of policy evaluation algorithms with improved exploration. We also note that we can use the bound of Theorem 2.2 in conjunction with TD(0)-type algorithms. In particular, in Example 5.2 we will 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.

4.1.2. Average cost and stochastic shortest path (SSP) problems. In the average cost case (similarly for SSP), x^* is the differential cost (also called the bias vector) of the policy to be evaluated, and x^* is orthogonal to e. To simplify the discussion, let us assume that the approximation subspace is orthogonal to e. Let ξ be the invariant distribution of the Markov chain. The mapping $\Pi A = \Pi P^{(\alpha,\lambda)}$ with $\alpha = 1, \lambda \in [0, 1)$ is nonexpansive with respect to $\|\cdot\|_\xi$, and the error bound corresponding to the bound (4), as given by Tsitsiklis and Van Roy [18], is

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

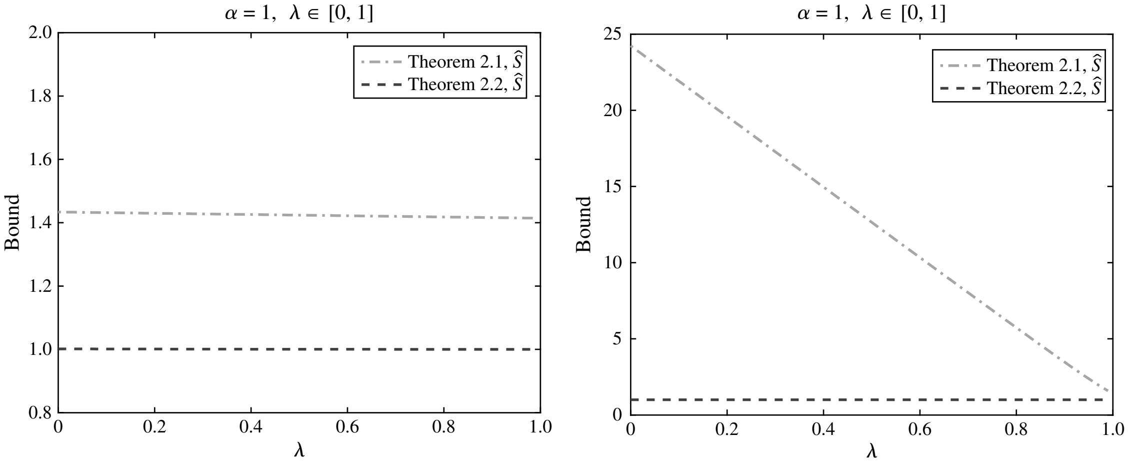
for some $\alpha_\lambda < 1$, with the property that $\alpha_\lambda \to 0$ as $\lambda \to 1$. In fact, α_λ is the modulus of contraction of some mapping of the form $\rho I + (1-\rho)\Pi A$, $\rho \in [0, 1)$, i.e., a damped version of ΠA, which attains the minimal modulus of contraction among all damped versions. The convergence of α_λ to zero as λ approaches 1 reflects the fact that with λ approaching 1, ΠA converges to the zero matrix (as A converges to the rank-one matrix $e\xi'$). However, the exact value of α_λ is usually unknown, so this bound provides no quantitative information.

Figure 6 shows the bounds of Theorems 2.1 and 2.2, with the term $\|\Pi A\|_\xi$ in the bound of Theorem 2.1 further upper-bounded by 1. Notice that as $\lambda \to 1$, the bound of Theorem 2.1 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 discussed in §2.2.2. Notice also in Figure 6(b) that the bound of Theorem 2.1 is affected by the relation of the approximation subspace to the eigenspace of A associated with eigenvalues that are close to 1, similar to the discounted case. By contrast, the bound of Theorem 2.2 performs well.

4.1.3. 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 cost. For simplicity of discussion, we consider the discounted case. The optimal cost function x^* is the unique solution of the nonlinear Bellman equation,

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

where g is the vector of one-stage costs associated with continuation, c is the vector of one-stage costs associated with stopping, and the minimization in $\min\{c, x\}$ is component-wise.



(a) Theorem 2.1 vs. Theorem 2.2

(b) Theorem 2.1 vs. Theorem 2.2

FIGURE 6. Comparison of error bounds for average cost problems with randomly generated Markov chains.
Notes. The setup is the same as 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) (Tsitsiklis and Van Roy [19], Choi and Van Roy [7], Yu and Bertsekas [21, 22]) solve the projected Bellman equation,

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

which is also nonlinear and has a unique solution $\bar{x}$ due to the contraction property of the mapping $\alpha\Pi P \min\{c, \cdot\}$ with respect to $\|\cdot\|_\xi$ (Tsitsiklis and Van Roy [19]). There is an error bound analogous to the bound (4): $\|\bar{x} - x^*\|_\xi \le (1/\sqrt{1-\alpha^2})\|x^* - \Pi x^*\|_\xi$ (Tsitsiklis and Van Roy [19]), and such error bound is also useful in bounding the performance of suboptimal policies constructed based on $\bar{x}$ (Van Roy [20]).

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

$$\bar{x} = \Pi g + \alpha \Pi P \min\{c, \bar{x}\} = \Pi g + \alpha \Pi P (\mathrm{I} - \mathrm{I}_{\bar{x}}) c + \alpha \Pi P \mathrm{I}_{\bar{x}} \bar{x}, \tag{54}$$

where $\mathrm{I}_{\bar{x}}$ is an $n \times n$ diagonal matrix with its ith diagonal entry defined by

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

We consider the linear equation

$$x = g + \alpha P (\mathrm{I} - \mathrm{I}_{\bar{x}}) c + \alpha P \mathrm{I}_{\bar{x}} x, \tag{55}$$

and its projected form

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

Both equations have a unique solution because $\alpha \Pi P \mathrm{I}_{\bar{x}}$ is a contraction mapping with respect to $\|\cdot\|_\xi$. The solution of the projected equation is $\bar{x}$, as can be seen by comparing the equation with Equation (54). Let $\hat{x}$ be the solution of Equation (55). If $\bar{x} = x^*$, then $\hat{x} = \bar{x} = x^*$, so the difference $\hat{x} - \bar{x}$ provides some information about the quality of $\bar{x}$ as an approximation of x^*. We can apply our error bounds with $A = \alpha P \mathrm{I}_{\bar{x}}$ to bound $\hat{x} - \bar{x}$, once $\bar{x}$ is computed. Estimating the matrices in the bounds using simulation data is no more complicated than in the policy evaluation case where there is no matrix $\mathrm{I}_{\bar{x}}$ involved in A, because $\mathrm{I}_{\bar{x}}$ is simply a diagonal matrix with 1s or 0s on its diagonal. Thus our error bounds can provide supplementary information about the approximation quality, in addition to the bounds based on the contraction property (Tsitsiklis and Van Roy [19], Van Roy [20]).

4.2. Large general systems of linear equations. For solving large general systems of linear equations using the projected equation approach (Bertsekas and Yu [4]), the bound of Theorem 2.2 can be computed in a straightforward way, as shown in Example 5.1 in §5. Theorem 2.2 is not only much sharper than Theorem 2.1 for this case but also more convenient, because it does not require the knowledge of $\|A\|_\xi$.

5. Estimating the low-dimensional matrices in the bounds.

We consider estimating the $k \times k$ matrices involved in the bounds of Theorems 2.1, 2.2 by simulation—if instead of using simulation, products of $k \times n$ and $n \times n$ matrices can be computed directly, then the calculation could be done directly with common matrix algebra.

The estimation of B and M using simulation has been well explained in the literature (see, e.g., Boyan [5], Nedić and Bertsekas [13], Bertsekas and Yu [4]), so we discuss here how to estimate the matrix R in Theorem 2.2:

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

The bounds in §3 involve other small-size matrices, which have similar forms to the matrices B, M, or R and can therefore be estimated using the same procedures for estimating the latter matrices, as explained in §3.

We demonstrate the estimation of R in two examples, one related to solving general linear equations and the other to policy evaluation in MDP, both with the TD(0)-type methods. For the TD(λ) case with $\lambda > 0$, which differs from TD(0) in that A itself is a summation of infinitely many matrices [cf. Equation (50) in the preceding section], we do not yet have an efficient way of estimating R.

Our methods for estimating R are based on a common procedure. Let $\phi(i)$ be the k-dimensional vector whose transpose $\phi(i)'$ is the ith row of Φ. Denote an entry of A and ξ by a_{ij} and ξ_i, respectively. We first express R as a certain summation of the $k \times k$ matrices $\phi(i)\phi(j)'$, $1 \le i, j \le n$, for instance,

$$R = \sum_{1 \le i, l, j \le n} (a_{il} a_{jl}) \cdot \frac{\xi_i \xi_j}{\xi_l} \cdot \phi(i)\phi(j)'. \tag{56}$$

Guided by such an expression, we generate a sequence of triple indices (i_t, j_t, l_t), $t \ge 0$, according to some probability distribution, and we choose proper weights $w_t(i_t, j_t, l_t)$ so that for all (i, j, l), we have the following match between the weighted long-run averages and the respective coefficients in the summation in Equation (56):

$$\frac{1}{t+1} \sum_{m=0}^{t} w_m(i_m, j_m, l_m)\delta[(i_m, j_m, l_m) = (i, j, l)] \;\longrightarrow\; (a_{il} a_{jl}) \cdot \frac{\xi_i \xi_j}{\xi_l}, \quad \text{as } t \to \infty, \text{ with probability 1}.$$

Here $\delta[\cdots]$ denotes the indicator for the event given in the brackets. As a consequence, the estimate R_t defined as the sample average

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

converges to R as $t \to \infty$ with probability 1.

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 because the normalization constant will be canceled in the product defining the matrix G_2 in the bound.

Example 5.1. Suppose both ξ and A are known explicitly. This case is relevant to the application of solving general linear equations, in which we know explicitly the entries of A, and we might want to choose a particular projection norm, for instance, the standard Euclidean norm (all entries of ξ being equal). We express R as the summation given in Equation (56) and generate a sequence of triple indices (i_t, j_t, l_t) as follows. We generate a sequence $(l_0, l_1, \ldots)$ such that the empirical frequencies of $1, 2, \ldots, n$ in this sequence converge to ξ. Given l_t, we generate two mutually independent transitions (l_t, i_t) and (l_t, j_t) according to a certain transition probability matrix P with $p_{ij} \ne 0$ whenever $a_{ji} \ne 0$. We then define R_t by

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

where t is a suitably large number. It is easy to see that with probability 1, for each (i, j, l), as $t \to \infty$,

$$\frac{1}{t+1} \sum_{m=0}^{t} \left(\frac{a_{i_m l_m}}{p_{l_m i_m}} \cdot \frac{a_{j_m l_m}}{p_{l_m j_m}} \right) \cdot \frac{\xi_{i_m} \xi_{j_m}}{\xi_{l_m}^2} \cdot \delta[(i_m, j_m, l_m) = (i, j, l)]$$

converges to

$$\xi_l p_{li} p_{lj} \cdot \left(\frac{a_{il}}{p_{li}} \cdot \frac{a_{jl}}{p_{lj}} \right) \cdot \frac{\xi_i \xi_j}{\xi_l^2} = (a_{il} a_{jl}) \cdot \frac{\xi_i \xi_j}{\xi_l}.$$

Therefore, comparing R_t with the expression of R in Equation (56), we have $R_t \to R$ as $t \to \infty$ with probability 1. We approximate R by the symmetrized matrix $(R_t + R_t')/2$. In the special case where $\xi_i = 1/n$ for all i, R reduces to $(1/n)\Phi' A A' \Phi$, the indices l_t can be generated independently with the uniform distribution, and the ratio $\xi_{i_m} \xi_{j_m} / \xi_{l_m}^2$ in R_t reduces to 1. □

Example 5.2. This example is relevant to the MDP applications, in particular, evaluating the cost or Q-factors (i.e., costs associated with state-action pairs) of a policy using TD(0)-like algorithms, with and without exploration enhancements. Suppose we do not know explicitly ξ and A. Moreover, suppose the ratios $\beta_{ij} = a_{ij}/p_{ij}$ are known for a certain transition matrix P with $p_{ij} \ne 0$ whenever $a_{ij} \ne 0$, and that 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 to generate transitions according to P. In the context of policy evaluation in MDP, P is the transition matrix of the Markov chain induced by a policy that is possibly different from the policy to be evaluated—the case with exploration; in the case without exploration, the two policies are the same, and $P = \beta A$ for some positive constant β. The estimation problem we have here is to estimate R using a trajectory $(i_0, i_1, \ldots)$ of the Markov chain associated with P.

Following the common procedure for estimating R outlined at the beginning of this section, we first express R as

$$R = \sum_{1\le i,\, l,\, j\le n} (\beta_{il}\beta_{jl}) \cdot \left(\xi_i p_{il} \cdot \frac{p_{jl}\xi_j}{\xi_l}\right) \cdot \phi(i)\phi(j)'. \tag{57}$$

The ratios β_{il}, β_{jl} are known under our assumption. What we need to approximate by properly weighted sample averages are the quantities $\xi_i p_{il}$ and $(p_{jl}\xi_j)/\xi_l$. For all (i, l), $\xi_i p_{il}$ are simply the joint probabilities of two consecutive states of the Markov chain when the chain is in equilibrium, so they can be approximated by the empirical frequencies of transitions $(i_{m-1} = i, i_m = l)$ occurred in the trajectory $(i_0, i_1, \ldots)$. The quantity $(p_{jl}\xi_j)/\xi_l$ equals *the conditional probability of the previous state being j given the current state being l* when the Markov chain is in equilibrium, so it can be approximated by the empirical frequency of the occurrence of $(i_{m-1} = j, i_m = l)$ among all transitions in the trajectory that end at l. This shows that we can approximate R as follows. At time t and given $i_{t+1} = l$, we draw one sample $(\hat{j}, l)$ uniformly from the set of past transitions that end at the state i_{t+1}, namely the set $\{(i_{t_k-1}, i_{t_k}) \mid i_{t_k} = l, t_k \le t+1\}$, and we let $j_t = \hat{j}$. (It also works to simply let j_t be the state immediately preceding the most recent visit to l prior to time $t+1$.) We then define R_t by

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

By the ergodicity of the Markov chain and the preceding discussion, it is clear that for each (i, j, l), as $t \to \infty$,

$$\frac{1}{t+1}\sum_{m=0}^{t} (\beta_{i_m i_{m+1}}\beta_{j_m i_{m+1}}) \cdot \delta[(i_m, j_m, i_{m+1}) = (i, j, l)] \;\longrightarrow\; (\beta_{il}\beta_{jl}) \cdot \left(\xi_i p_{il} \cdot \frac{p_{jl}\xi_j}{\xi_l}\right)$$

with probability 1, so, by comparing R_t with the expression of R in Equation (57), the convergence of R_t to R as $t \to \infty$ with probability 1 is evident. We approximate R by the symmetrized matrix $(R_t + R_t')/2$.

It can be seen that generating "backward" transitions $(\hat{j}, l)$ to a given state/index l according to the steady-state conditional distribution increases the memory and computational requirements, because the past history of the simulation must be stored and searched. If the Markov chain is reversible (i.e., $\xi_j p_{jl} = \xi_l p_{lj}$ for all j, l), then we can omit the step of generating backward transitions by using the reversibility property. The reason is that when the chain is in equilibrium, a forward transition (i_{m+1}, i_{m+2}) from i_{m+1} automatically gives us such a backward transition: the probability of i_{m+2} given i_{m+1} is $p_{i_{m+1}i_{m+2}} = (\xi_{i_{m+2}} p_{i_{m+2}i_{m+1}})/\xi_{i_{m+1}}$. Correspondingly, the estimation procedure can be substantially simplified by letting $j_m = i_{m+2}$ in R_t.

When the weight vector ξ is given explicitly and need not be the invariant distribution of P, a variant of the above procedure is possible. In this case, we can approximate the invariant distribution of P, denoted by $\bar{\kappa}$, using the empirical frequencies of $1, 2, \ldots, n$ in the trajectory $(i_0, i_1, \ldots)$ up to time t. Let $\kappa_{t,\, i}$ be the empirical frequency of the occurrence of $i_m = i$ in the trajectory up to time t. We define R_t by

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

and we approximate R by the symmetrized matrix $(R_t + R_t')/2$. The reasoning on the convergence of R_t to R as $t \to \infty$ is similar to the case discussed earlier. By the ergodicity of the Markov chain, for each (i, j, l), as $t \to \infty$,

$$\frac{1}{t+1}\sum_{m=0}^{t} (\beta_{i_m i_{m+1}}\beta_{j_m i_{m+1}}) \cdot \left(\frac{\xi_{i_m}\xi_{j_m}}{\xi_{i_{m+1}}} \cdot \frac{\kappa_{t,\, i_{m+1}}}{\kappa_{t,\, i_m}\kappa_{t,\, j_m}}\right) \cdot \delta[(i_m, j_m, i_{m+1}) = (i, j, l)]$$

converges to

$$(\beta_{il}\beta_{jl}) \cdot \left(\bar{\kappa}_i p_{il} \cdot \frac{p_{jl}\bar{\kappa}_j}{\bar{\kappa}_l} \cdot \frac{\xi_i\xi_j}{\xi_l} \cdot \frac{\bar{\kappa}_l}{\bar{\kappa}_i\bar{\kappa}_j}\right) = (\beta_{il}\beta_{jl}) \cdot \left(\xi_i p_{il} \cdot \frac{p_{jl}\xi_j}{\xi_l}\right),$$

with probability 1. Comparing R_t with Equation (57), the convergence of R_t to R as $t \to \infty$ follows. □

We summarize the discussion in the two preceding examples. When ξ and A are known, it is straightforward to estimate R. In the MDP context, where ξ or A are unknown, there is a memory and computation issue when simulating backward transitions according to the steady-state conditional distribution, as well as when a desirable weight vector ξ does not match the frequencies of the indices generated by simulation. The procedures given above do not adapt easily to the case where A itself is a summation of infinitely many matrices, as in the multistep projected Bellman equations solved by TD(λ) with $\lambda > 0$.

6. Conclusion. We have derived new data-dependent computable error bounds for the approximate solution of large linear systems using the Galerkin/projected equation approach, as well as a least-squares equation error approach. The bounds hold for both contraction and noncontraction mappings. Their applicability for noncontraction mappings is not only useful for approximating solutions of general linear equations but is also useful in the context of MDP when using exploration to evaluate the cost of policies. Furthermore, in the context of MDP, our bounds can be used in performance bounds for approximate policy iteration, such as those of Munos [12].

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 indicates that the simulation noise tends to increase as λ increases, so there is strong motivation to use small values of λ as long as the amplification ratio is close to 1. On the other hand, the use of small values of λ might result in unacceptably large amplification ratio, as demonstrated analytically by examples given in Bertsekas [1]; see also Bertsekas and Tsitsiklis [3, Example 6.5, pp. 288–289]. Unfortunately, the bounds (3), (4) 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 noncontraction context, in conjunction with simulation-based TD(λ)-like algorithms that have been developed in our recent paper (Bertsekas and Yu [4]). There might 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.

Acknowledgments. The authors thank Professors Remi Munos and Bruno Scherrer for helpful discussions and the two reviewers for providing many comments that helped improve the presentation of the paper. 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. A preliminary version of this paper appeared in *Proceedings of the* 8*th European Workshop on Reinforcement Learning* (EWRL 2008), *Lecture Notes in Computer Science*, Vol. 5323. Springer, Berlin, 253–267.

References

[1] Bertsekas, D. P. 1995. A counterexample to temporal differences learning. *Neural Comput.* **7** 270–279.
[2] Bertsekas, D. P. 2007. *Dynamic Programming and Optimal Control*, 3rd ed., Vol. II. Athena Scientific, Belmont, MA.
[3] Bertsekas, D. P., J. N. Tsitsiklis. 1996. *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA.
[4] Bertsekas, D. P., H. Yu. 2009. Projected equation methods for approximate solution of large linear systems. *J. Comput. Appl. Math.* **227**(1) 27–50.
[5] Boyan, J. A. 1999. Least-squares temporal difference learning. *Proc. 16th Internat. Conf. Machine Learn.* Morgan Kaufmann, San Francisco, 49–56.
[6] Bradtke, S. J., A. G. Barto. 1996. Linear least-squares algorithms for temporal difference learning. *Machine Learn.* **22**(2) 33–57.
[7] Choi, D. S., B. Van Roy. 2006. A generalized Kalman filter for fixed point approximation and efficient temporal-difference learning. *Discrete Event Dynam. Systems* **16**(2) 207–239.
[8] Horn, R. A., C. R. Johnson. 1985. *Matrix Analysis*. Cambridge University Press, Cambridge, UK.
[9] Konda, V. R. 2002. Actor-critic algorithms. Thesis, Massachusetts Institute of Technology, Cambridge.
[10] Konda, V. R., J. N. Tsitsiklis. 2003. Actor-critic algorithms. *SIAM J. Control Optim.* **42**(4) 1143–1166.
[11] Krasnose'skii, M. A., G. M. Vainikko, P. P. Zabreiko, Ya. B. Rutitskii, V. Ya. Stetsenko. 1972. *Approximate Solution of Operator Equations*. Wolters-Noordhoff Publishing, Groningen, The Netherlands.
[12] Munos, R. 2003. Error bounds for approximate policy iteration. *Proc. 20th Int. Conf. Machine Learning*, AUAI Press, Washington, DC, 560–567.
[13] Nedić, A., D. P. Bertsekas. 2003. Least squares policy evaluation algorithms with linear function approximation. *Discrete Event Dynam. Systems* **13** 79–110.
[14] Sutton, R. S. 1988. Learning to predict by the methods of temporal differences. *Machine Learn.* **3** 9–44.
[15] Sutton, R. S., A. G. Barto. 1998. *Reinforcement Learning*. MIT Press, Cambridge, MA.
[16] Szyld, D. B. 2006. The many proofs of an identity on the norm of oblique projections. *Numer. Algorithms* **42**(3–4) 309–323.
[17] Tsitsiklis, J. N., B. Van Roy. 1997. An analysis of temporal-difference learning with function approximation. *IEEE Trans. Automatic Control* **42**(5) 674–690.
[18] Tsitsiklis, J. N., B. Van Roy. 1999a. Average cost temporal-difference learning. *Automatica* **35**(11) 1799–1808.
[19] Tsitsiklis, J. N., B. Van Roy. 1999b. Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing financial derivatives. *IEEE Trans. Automatic Control* **44**(10) 1840–1851.
[20] Van Roy, B. 2007. On regression-based stopping times. *Discrete Event Dynam. Systems*, ePub ahead of print February 7, 2009, http://www.springerlink.com/content/831433v414640767/.
[21] Yu, H., D. P. Bertsekas. 2006. A least squares Q-learning algorithm for optimal stopping problems. LIDS Technical Report 2731, Massachusetts Institute of Technology, Cambridge.
[22] Yu, H., D. P. Bertsekas. 2007. Q-learning algorithms for optimal stopping based on least squares. *Proc. Eur. Control Conf., Kos, Greece*, 2368–2375.