Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Optimal Oracle Inequalities for Projected Fixed-Point Equations, with Applications to Policy Evaluation

Wenlong Mou, Ashwin Pananjady, Martin J. Wainwright

To cite this article:
Wenlong Mou, Ashwin Pananjady, Martin J. Wainwright (2023) Optimal Oracle Inequalities for Projected Fixed-Point Equations, with Applications to Policy Evaluation. Mathematics of Operations Research 48(4):2308-2336. https://doi.org/10.1287/moor.2022.1341

https://pubsonline.informs.org/journal/moor

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 48, No. 4, November 2023, pp. 2308–2336
ISSN 0364-765X (print), ISSN 1526-5471 (online)

Optimal Oracle Inequalities for Projected Fixed-Point Equations, with Applications to Policy Evaluation

Wenlong Mou,[a,*] Ashwin Pananjady,[b] Martin J. Wainwright[a,c]

[a] University of California, Berkeley, Berkeley, California 94720; [b] Georgia Institute of Technology, Atlanta, Georgia 30332; [c] Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

*Corresponding author

Contact: wmou@eecs.berkeley.edu, https://orcid.org/0000-0001-9721-4213 (WM); ashwinpm@gatech.edu, https://orcid.org/0000-0003-0824-9815 (AP); wainwrigwork@gmail.com, https://orcid.org/0000-0002-8760-2236 (MJW)

Received: February 20, 2021
Revised: September 5, 2022
Accepted: September 28, 2022
Published Online in Articles in Advance: December 28, 2022

MSC2020 Subject Classifications: Primary: 62J05, 62M05, 65M15, 90C40, 93E24

https://doi.org/10.1287/moor.2022.1341

Abstract. Linear fixed-point equations in Hilbert spaces arise in a variety of settings, including reinforcement learning, and computational methods for solving differential and integral equations. We study methods that use a collection of random observations to compute approximate solutions by searching over a known low-dimensional subspace of the Hilbert space. First, we prove an instance-dependent upper bound on the mean-squared error for a linear stochastic approximation scheme that exploits Polyak–Ruppert averaging. This bound consists of two terms: an approximation error term with an instance-dependent approximation factor and a statistical error term that captures the instance-specific complexity of the noise when projected onto the low-dimensional subspace. Using information-theoretic methods, we also establish lower bounds showing that both of these terms cannot be improved, again in an instance-dependent sense. A concrete consequence of our characterization is that the optimal approximation factor in this problem can be much larger than a universal constant. We show how our results precisely characterize the error of a class of temporal difference learning methods for the policy evaluation problem with linear function approximation, establishing their optimality.

Funding: This work was partially supported by grants from the Office of Naval Research [Grant DOD-ONR-N00014-18-1-2640] and the National Science Foundation (NSF) [NSF-IIS Grant 1909365, NSF-DMS Grant 2015454, and NSF-CCF Grant 1955450] to M. J. Wainwright. Part of this work was performed when A. Pananjady was visiting the Simons Institute for the Theory of Computing, where he was supported by a Swiss Re Research Fellowship.

Supplemental Material: The online supplementary file is available at https://doi.org/10.1287/moor.2022.1341.

Keywords: linear fixed-point equations • Galerkin methods • temporal-difference methods • stochastic approximation • local minimax lower bounds

1. Introduction

Linear fixed-point equations over a Hilbert space, with the Euclidean space being an important special case, arise in various contexts. Such fixed-point equations take different names in different domains, including estimating equations, Bellman equations, Poisson equations, and inverse systems (Bertsekas [5], Krasnosel'skii et al. [31], Wooldridge [69]). More specifically, given a Hilbert space $\mathbb{X}$, we consider a fixed-point equation of the form

$$v = Lv + b, \tag{1}$$

where b is some member of the Hilbert space, and L is a linear operator mapping $\mathbb{X}$ to itself.

When the Hilbert space is infinite-dimensional—or has a finite but very large dimension D—it is common to seek approximate solutions to Equation (1). A standard approach is to choose a subspace $\mathbb{S}$ of the Hilbert space, of dimension $d \ll D$, and to search for solutions within this subspace. In particular, letting $\Pi_{\mathbb{S}}$ denote the orthogonal projection onto this subspace, various methods seek (approximate) solutions to the *projected fixed-point equation*:

$$v = \Pi_{\mathbb{S}}(Lv + b). \tag{2}$$

In order to set the stage, let us consider some generic examples that illustrate the projected fixed-point equation (2). We eschew a fully rigorous exposition at this stage, deferring technical details and specific examples to Section 2.2.

Example 1 (Galerkin Methods for Differential Equations). Let $\mathbb{X}$ be a Hilbert space of suitably differentiable functions, and let A be a linear differential operator of order k-say, of the form $A(v) = \omega_0 v + \sum_{j=1}^{k} \omega_j v^{(j)}$, where $v^{(j)}$ denotes the jth-order derivative of the function $v \in \mathbb{X}$. Given a function $b \in \mathbb{X}$, suppose that we are interested in solving the differential equation $A(v) = b$. This represents a particular case of our fixed-point equation with $L = I - A$.

Let $\mathbb{S}$ be a finite-dimensional subspace of $\mathbb{X}$-say, spanned by a set of basis functions $\{\phi_j\}_{j=1}^{d}$. A Galerkin method constructs an approximate solution to the differential equation $A(v) = b$ by solving the projected fixed-point equation (2) over a subspace of this type. Concretely, any function $v \in \mathbb{S}$ has a representation of the form $v = \sum_{j=1}^{d} \vartheta_j \phi_j$ for some weight vector $\vartheta \in \mathbb{R}^d$. Applying the operator A to any such function yields the residual $A(v) = \sum_{j=1}^{d} \vartheta_j A(\phi_j)$, and the Galerkin method chooses the weight vector $\vartheta \in \mathbb{R}^d$ such that v satisfies the equation $v = \Pi_{\mathbb{S}}((I - A)v + b)$. In Section 2.2.2, we describe in detail a specific version of the Galerkin method as applied to a second-order differential equation that underlies the so-called *elliptic boundary value problem*.

Example 2 (Instrumental Variable Methods for Nonparametric Regression). Let $\mathbb{X}$ denote a suitably constrained space of square-integrable functions mapping $\mathbb{R}^p \to \mathbb{R}$, and suppose that we have a regression model of the form $Y = f^*(X) + \epsilon$. Here, X is a random vector of covariates taking values in $\mathbb{R}^p$, the pair (Y, ϵ) denotes scalar random variables, and $f^* \in \mathbb{X}$ denotes an unknown function of interest. For a discussion of the existence and uniqueness of the various objects in this model, see Darolles et al. [20].

In the classical setup of nonparametric regression, it is assumed that $\mathbb{E}[\epsilon \mid X] = 0$, an assumption that can be violated. Instead, suppose that we have a vector of *instrumental variables* $Z \in \mathbb{R}^p$ such that $\mathbb{E}[\epsilon|Z] = 0$. Now let $T : \mathbb{X} \to \mathbb{X}$ denote a linear operator given by $T(f) = \mathbb{E}[f(X)|Z]$, and denote by $r = \mathbb{E}[Y|Z]$ a point in $\mathbb{X}$. Instrumental variable (IV) approaches to estimating f^* are based on the equality

$$\mathbb{E}[Y - f^*(X) \mid Z] = r - T(f^*) = 0, \tag{3}$$

which is a linear fixed-point relation of the form (1) with $L = I - T$ and $b = r$.

Now let $\{\phi_j\}_{j \geq 1}$ be an orthonormal basis of $\mathbb{X}$, and let $\mathbb{S}$ denote the subspace spanned by the first d such eigenfunctions. Then each function $f \in \mathbb{S}$ can be represented as $f = \sum_{j=1}^{d} \vartheta_j \phi_j$, and approximate solutions to the fixed-point equation (3) may be obtained via solving a projected variant (2) (i.e., the equation $f = \Pi_{\mathbb{S}}((I - T)f + r)$).

A specific example of an IV method is the class of temporal difference methods for policy evaluation, introduced and discussed in detail in Section 2.2.3.

In particular instantiations of both Examples 1 and 2, it is typical for the ambient dimension D to be very large (if not infinite) and for us to only have sample access to the pair (L, b). This paper treats the setting in which n observations $\{(L_i, b_i)\}_{i=1}^{n}$ are drawn independent and identically distributed (i.i.d.) from some distribution with mean (L, b). Letting v^* denote the solution to the fixed-point equation (1), our goal is to use these observations to produce an estimate $\hat{v}_n$ of v^* that satisfies an *oracle inequality* of the form

$$\mathbb{E}\|\hat{v}_n - v^*\|^2 \leq \alpha \cdot \inf_{v \in \mathbb{S}} \|v - v^*\|^2 + \varepsilon_n. \tag{4}$$

Here, we use $\|\cdot\|$ to denote the Hilbert norm associated with $\mathbb{X}$. The three terms appearing on the right-hand side of Inequality (4) all have concrete interpretations. The term

$$\mathcal{A}(\mathbb{S}, v^*) := \inf_{v \in \mathbb{S}} \|v - v^*\|^2 \tag{5}$$

defines the *approximation error*; this is the error incurred by an oracle procedure that knows the fixed point v^* in advance and aims to output the best approximation to v^* within the subspace $\mathbb{S}$. The term α is the *approximation factor*, which indicates how poorly the estimator $\hat{v}_n$ performs at carrying out the aforementioned approximation; note that $\alpha \geq 1$ by definition, and it is most desirable for α to be as small as possible. The final term ε_n is a proxy for the *statistical error* incurred as a result of our stochastic observation model; indeed, one expects that as the sample size n goes to infinity, this error should tend to zero for any reasonable estimator, indicating consistent estimation when $v^* \in \mathbb{S}$. More generally, we would like our estimator to also have as small a statistical error as possible in terms of the other parameters that define the problem instance.

In an ideal world, both desiderata hold simultaneously: the approximation factor should be as close to 1 as possible while the statistical error stays as small as possible. As we discuss shortly, such a "best-of-both-worlds" guarantee can indeed be obtained in many canonical problems, and "sharp" oracle inequalities—meaning ones in which the approximation factor is equal to 1—are known (Dalalyan and Salmon [18], Rakhlin et al. [50]). On

the other hand, such oracle equalities with unit factors are not known for the fixed-point equation (1). Tsitsiklis and Van Roy [60] show that if the operator L is $\gamma_{\max}$-contractive in the norm $\|\cdot\|$, then the (deterministic) solution $\overline{v}$ to the projected fixed-point equation (2) satisfies the bound

$$\|\overline{v} - v^*\|^2 \leq \frac{1}{1-\gamma_{\max}^2} \inf_{v \in \mathbb{S}} \|v - v^*\|^2. \tag{6}$$

Because $\gamma_{\max}$ can be arbitrarily close to 1, the prefactor in the bound (6) can be much larger than 1, in contrast to so-called sharp oracle inequalities for nonparametric regression. One motivating question for our work is whether this bound can be improved and, if so, to what extent.[1]

Our work is also driven by the complementary question of whether a sharp bound can be obtained on the statistical error of an estimator that, unlike $\overline{v}$, has access only to the samples $\{(L_i, b_i)\}_{i=1}^n$. In particular, we would like the statistical error ε_n to depend on some notion of complexity within the subspace $\mathbb{S}$ and *not* on the ambient space. Recent work by Bhandari et al. [8] provides worst-case bounds on the statistical error of a stochastic approximation scheme, showing that the *parametric rate* $\epsilon_n \lesssim d/n$ is attainable. In this paper, we study how to derive a more fine-grained bound on the statistical error that reflects the practical performance of the algorithm and depends optimally on the geometry of our problem instance.

1.1. Contributions and Organization

The main contribution of this paper is to resolve both of the aforementioned questions, in particular by deriving upper bounds and information-theoretic lower bounds on both the approximation factor and statistical error that are *instance dependent*. On one hand, these bounds demonstrate that, in general, it is not possible to obtain an oracle inequality with a prefactor equal to 1 but that there are many settings in which the optimal approximation factor is much smaller than what is suggested by the worst-case bound (6). We also derive a significantly sharper bound on the statistical error of a stochastic approximation scheme that is instance optimal in a precise sense. In more detail, the contributions of this paper include the following:

- Theorem 1 establishes an instance-dependent upper bound of the form (4) for the Polyak–Ruppert averaged stochastic approximation estimator, whose approximation factor α depends in a precise way on the projection of the operator L onto the subspace $\mathbb{S}$, and the statistical error ϵ_n matches the Cramér–Rao lower bound for the instance within the subspace.
- In Theorem 2, we prove an information-theoretic lower bound on the approximation factor. It is a local analysis, in that the bound depends critically on the projection of the population-level operator. This lower bound certifies that the approximation factor attained by our estimator is optimal. To the best of our knowledge, this is also the first instance of an optimal oracle inequality with a nonconstant and problem-dependent approximation factor.
- In Theorem 3, we establish via a Bayesian Cramér–Rao lower bound that the leading statistical error term for our estimator is also optimal in an instance-dependent sense.
- In Section 4, we derive specific consequences of our results for several examples, including the problem of Galerkin approximation in second-order elliptic equations, as well as temporal difference methods for policy evaluation with linear function approximation. A particular consequence of our results shows that in a minimax sense, the approximation factor (6) is optimal for a policy evaluation with linear function approximation (cf. Proposition 1).

The remainder of this paper is organized as follows. Section 1.2 contains a detailed discussion of related work. We introduce formal background and specific examples in Section 2. Our main results under the general model of projected fixed-point equations are introduced and discussed in Section 3. We then specialize these results to our examples in Section 4, deriving several concrete corollaries for Galerkin methods and temporal difference methods. Our proofs are presented in Section 5, and technical results are relegated to the appendix in the online supplementary file.

1.2. Related Work

Our paper touches on various lines of related work, including stochastic approximation and its application to reinforcement learning, projected linear equation methods, as well as oracle inequalities for statistical estimation. Let us provide a brief discussion of these connections here.

1.2.1. Stochastic Approximation. Stochastic approximation algorithms for both linear and nonlinear fixed-point equations play a central role in large-scale machine learning and statistics (Lai [32], Nemirovski et al. [43], Robbins and Sutton [52]). See the books by Benveniste et al. [4] and Borkar [9] for a comprehensive survey of the classical methods of analysis. In the seminal work of Polyak, Ruppert, and Juditsky (Polyak [46], Polyak and Juditsky [47], Ruppert [54]), the authors proposed taking the average of the stochastic approximation iterates,

which stabilizes the algorithm and ensures a Gaussian limiting distribution. In fact, the averaged iterates are known to be asymptotically optimal in a local minimax sense (Duchi and Ruan [21]). Nonasymptotic guarantees matching this asymptotic behavior have also been established for other forms of stochastic approximation, as well as variance-reduced variants thereof (Khamaru et al. [26], Li et al. [36], Mou et al. [40], Moulines and Bach [41]).

Stochastic approximation is also a fundamental building block for reinforcement learning algorithms, wherein the method is used to produce an iterative, online solution to the Bellman equation from data; see the books by Szepesvári [59] and Bertsekas [7] for a survey. Such approaches include temporal difference (TD) methods (Sutton [57]) for the policy evaluation problem and the Q-learning algorithm (Watkins and Dayan [68]) for policy optimization. Variants of these algorithms also abound, including least squares temporal difference (LSTD) (Boyan [11]), state–action–reward–state–action (SARSA; Rummery and Niranjan [53]), actor-critic algorithms (Konda and Tsitsiklis [30]), and gradient TD methods (Sutton et al. [58]). The analysis of these methods has received significant attention in the literature, ranging from asymptotic guarantees (e.g., Bradtke and Barto [12], Tsitsiklis and Van Roy [60, 61]) to more fine-grained finite-sample bounds (e.g., Bhandari et al. [8], Lakshminarayanan and Szepesvári [33], Pananjady and Wainwright [45], Srikant and Ying [56], Wainwright [66, 67]). Our work contributes to this literature, because as a corollary of our general analysis, we are able to establish finite-sample upper bounds for temporal difference methods with Polyak–Ruppert averaging, as applied to the policy evaluation problem with linear function approximation.

1.2.2. Projected Methods for Linear Equations. In 1915, Galerkin [23] first proposed the method of approximating the solution to a linear partial differential equation (PDE) by solving the projected equation in a finite-dimensional subspace. This method later became a cornerstone of finite-element methods in numerical methods for PDEs; see the chapter by Fletcher [22] and the book by Brenner and Scott [13] for a comprehensive survey. A fundamental tool used in the analysis of Galerkin methods is Céa's lemma (Céa [16]); in this paper, we derive more general upper bounds on the approximation factor that capture this classical lemma as a special case. As mentioned before, in the specific context of reinforcement learning, projected linear equations were studied by Tsitsiklis and Van Roy [60], who first proved the upper bound (6) on the approximation factor under contractivity assumptions. These contraction-based bounds were further extended to the analysis of Q-learning in optimal stopping problems (Tsitsiklis and Van Roy [61]). The connection between the Galerkin method and TD methods was observed by Yu and Bertsekas [71] and Bertsekas [5], and the former paper provides an instance-dependent upper bound on the approximation factor. This analysis was later applied to Monte Carlo methods for solving linear inverse problems (Polydorides et al. [48, 49]).

The Bellman equation can be written in infinitely many equivalent ways—by using powers of the transition kernel and via the formalism of resolvents—leading to a continuous family of projected equations indexed by a scalar parameter λ (see, e.g., section 5.5 of Bertsekas [7]). Some of these forms can be specifically leveraged in other observation models; for instance, by observing the trajectory of the Markov chain instead of i.i.d. samples, it becomes possible to obtain unbiased observations for integer powers of the transition kernel. This makes it possible to efficiently estimate the solution to the projected linear equation for various values of λ, and it underlies the family of TD(λ) methods (Boyan [11], Sutton [57]). Indeed, Tsitsiklis and Van Roy [60] also showed that the worst-case approximation factor in Equation (6) can be improved by using larger values of λ. Based on this observation, a line of work has studied the trade-off between the approximation error and estimation measure in model selection for reinforcement learning problems (Bertsekas [6], Munos and Szepesvári [42], Scherrer [55], Van Roy [64]). Understanding precise trade-offs between the approximation and estimation error is crucial to model selection. However, unlike this body of work, our focus in the current paper is on studying the i.i.d. observation model; a detailed investigation into the Markov setting is an important direction for future work.

1.2.3. Oracle Inequalities. There is a large literature on misspecified statistical models and oracle inequalities (e.g., see the monographs by Massart [38] and Koltchinskii [29] for overviews). Oracle inequalities in the context of penalized empirical risk minimization (ERM) are quite well understood (e.g., Bartlett et al. [2], Koltchinskii [28], and Massart and Nédélec [39]). Typically, the resulting approximation factor is exactly 1 or arbitrarily close to 1, and the statistical error term depends on the localized Rademacher complexity or metric entropy of this function class. Aggregation methods have been developed in order to obtain *sharp* oracle inequalities with approximation factor exactly 1 (e.g., Bunea et al. [14], Dalalyan and Salmon [18], Rakhlin et al. [50], and Tsybakov [62]). Sharp oracle inequalities are now available in a variety of settings, including for sparse linear models (Bunea et al. [15]), density estimation (Dalalyan and Sebbar [19]), graphon estimation (Klopp et al. [27]), and shape-constrained estimation (Bellec [3]). As previously noted, our setting differs qualitatively from the ERM

setting in that, as shown in this paper, sharp oracle inequalities are no longer possible. There is another related line of work on oracle inequalities of density estimation. Yatracos [70] showed an oracle inequality with the non-standard approximation factor 3 and with a statistical error term depending on the metric entropy. This non-unit approximation factor was later shown to be optimal for the class of one-dimensional piecewise constant densities (Bousquet et al. [10], Chan et al. [17], Zhu et al. [72]). The approximation factor lower bound in these papers and that in our work both make use of the birthday paradox to establish information-theoretic lower bounds.

1.3. Notation

Here, we summarize some notation used throughout the paper. For a positive integer m, we define the set $[m] := \{1, 2, \ldots, m\}$. For any pair $(\mathbb{X}, \mathbb{Y})$ of real Hilbert spaces and a linear operator $A : \mathbb{X} \to \mathbb{Y}$, we denote by $A^* : \mathbb{Y} \to \mathbb{X}$ the adjoint operator of A, which, by definition, satisfies $\langle Ax, y\rangle = \langle x, A^*y\rangle$ for all $(x, y) \in \mathbb{X} \times \mathbb{Y}$. For a bounded linear operator A from $\mathbb{X}$ to $\mathbb{Y}$, we define its operator norm as $|||A|||_{\mathbb{X}\to\mathbb{Y}} := \sup_{x\in\mathbb{X}\setminus\{0\}} \frac{\|Ax\|_{\mathbb{Y}}}{\|x\|_{\mathbb{X}}}$. We use the shorthand notation $|||A|||_{\mathbb{X}}$ to denote its operator norm when A is a bounded linear operator mapping $\mathbb{X}$ to itself. When $\mathbb{X} = \mathbb{R}^{d_1}$ and $\mathbb{Y} = \mathbb{R}^{d_2}$ are finite-dimensional Euclidean spaces equipped with the standard inner product, we denote by $|||A|||_{\mathrm{op}}$ the operator norm in this case. We also use $\|\cdot\|_2$ to denote the standard Euclidean norm, in order to distinguish it from the Hilbert norm $\|\cdot\|$.

For a random object X, we use $\mathcal{L}(X)$ to denote its probability law. Given a vector $\mu \in \mathbb{R}^d$ and a positive semidefinite matrix $\Sigma \in \mathbb{R}^{d\times d}$, we use $\mathcal{N}(\mu, \Sigma)$ to denote the Gaussian distribution with mean μ and covariance Σ. We use $\mathcal{U}(\Omega)$ to denote the uniform distribution over a set Ω. Given a Polish space $\mathcal{S}$ and a positive measure μ associated to its Borel σ-algebra, for $p \in [1, +\infty)$, we define $\mathbb{L}^p(\mathcal{S}, \mu) := \{f : \mathcal{S} \to \mathbb{R}, \|f\|_{\mathbb{L}^p} := (\int_{\mathcal{S}} |f|^p d\mu)^{1/p} < +\infty\}$. When $\mathcal{S}$ is a subset of $\mathbb{R}^d$ and μ is the Lebesgue measure, we use the shorthand notation $\mathbb{L}^p(\mathcal{S})$. For a point $x \in \mathbb{R}^d$, we use δ_x to denote the Dirac δ-function at point x.

We use $\{e_j\}_{j=1}^d$ to denote the standard basis vectors in the Euclidean space $\mathbb{R}^d$ (i.e., e_i is a vector with a 1 in the ith coordinate and 0s elsewhere). For two matrices $A \in \mathbb{R}^{d_1\times d_2}$ and $B \in \mathbb{R}^{d_3\times d_4}$, we denote by $A \otimes B$ their Kronecker product, a $d_1d_3 \times d_2d_4$ real matrix. For symmetric matrices $A, B \in \mathbb{R}^{d\times d}$, we use $A \preceq B$ to denote the fact $B - A$ is a positive semidefinite matrix and use $A \prec B$ when $B - A$ is positive definite. For a positive integer d and indices $i, j \in [d]$, we denote by E_{ij} a $d \times d$ matrix with a 1 in the (i, j) position and 0s elsewhere. More generally, given a set $\mathcal{S}$ and $s_1, s_2, \in \mathcal{S}$, we define E_{s_1,s_2} to be the linear operator such that $E_{s_1,s_2} f(x) := f(s_2)\mathbf{1}_{x=s_1}$ for all $f : \mathcal{S} \to \mathbb{R}$.

2. Background

We begin by formulating the projected fixed-point problem more precisely in Section 2.1. Section 2.2 provides illustrations of this general setup with some concrete examples.

2.1. Problem Formulation

Consider a separable Hilbert space $\mathbb{X}$ with (possibly infinite) dimension D, equipped with the inner product $\langle\cdot, \cdot\rangle$. Let $\mathfrak{L}$ denote the set of all bounded linear operators mapping $\mathbb{X}$ to itself. Given one such operator $L \in \mathfrak{L}$ and some $b \in \mathbb{X}$, we consider the fixed-point relation $v = Lv + b$, as previously defined in Equation (1). We assume that the operator $I - L$ has a bounded inverse, which guarantees the existence and uniqueness of the fixed point satisfying Equation (1). We let v^* denote this unique solution.

As previously noted, in general, solving a fixed-point equation in the Hilbert space can be computationally challenging. Consequently, a natural approach is to seek approximations to the fixed point v^* based on searching over a finite-dimensional subspace of the full Hilbert space. More precisely, given some d-dimensional subspace $\mathbb{S}$ of $\mathbb{X}$, we seek to solve the projected fixed-point Equation (2).

2.1.1. Existence and Uniqueness of a Projected Fixed Point. For concreteness in analysis, we are interested in problems for which the projected fixed-point equation has a unique solution. Here, we provide a sufficient condition for such existence and uniqueness. In doing so, and for future reference, it is helpful to define some mappings between $\mathbb{X}$ and the subspace $\mathbb{S}$. Let us fix some orthogonal basis $\{\phi_j\}_{j\geq 1}$ of the full space $\mathbb{X}$ such that $\mathbb{S} = \text{span}\ \{\phi_1, \ldots, \phi_d\}$. In terms of this basis, we can define the projection operator $\Phi_d : \mathbb{X} \to \mathbb{R}^d$ via $\Phi_d(x) := (\langle x, \phi_j\rangle)_{j=1}^d$. The adjoint operator of Φ_d is a mapping from $\mathbb{R}^d$ to $\mathbb{X}$, given by

$$\Phi_d(v) := \sum_{j=1}^{d} v_j \phi_j. \tag{7}$$

Using these operators, we can define the *projected operator* associated with L—namely,

$$M := \Phi_d L \Phi_d^*. \tag{8}$$

Note that M is simply a d-dimensional matrix, one that describes the action of L on $\mathbb{S}$ according to the basis that we have chosen. As we will see in the main theorems, our results do not depend on the specific choice of the orthonormal basis, but it is convenient to use a given one, as we have done here.

Consider the quantity

$$\kappa(M) := \frac{1}{2}\lambda_{\max}(M + M^\top), \tag{9}$$

which corresponds to the maximal eigenvalue of the symmetrized version of M. One sufficient condition for there be a unique solution to the fixed-point equation (2) is the bound $\kappa(M) < 1$. When this bound holds, the matrix $(I_d - M)$ is invertible, and hence for any $b \in \mathbb{X}$, there is a unique solution $\overline{v}$ to the equation $v = \Pi_{\mathbb{S}}(Lv + b)$.

2.1.2. Stochastic Observation Model. As noted in the introduction, this paper focuses on an observation model in which we observe i.i.d. random pairs (L_i, b_i) for $i = 1, \ldots, n$ that are unbiased estimates of the pair (L, b) so that

$$\mathbb{E}[L_i] = L, \quad \text{and} \quad \mathbb{E}[b_i] = b. \tag{10}$$

In addition to this unbiasedness, we also assume that our observations satisfy a certain second-moment bound. A weaker version and a stronger version of this assumption are both considered.

Assumption 1A (Second-Moment Bound in Projected Space (Weak)). *There exist scalars $\sigma_L, \sigma_b > 0$ such that for any unit-norm vector $u \in \mathbb{S}$ and any basis vector in $\{\phi_j\}_{j=1}^d$ we have the bounds*

$$\mathbb{E}\langle \phi_j, (L_i - L)u\rangle^2 \le \sigma_L^2 \|u\|^2, \quad \text{and} \tag{11a}$$

$$\mathbb{E}\langle \phi_j, b_i - b\rangle^2 \le \sigma_b^2. \tag{11b}$$

Assumption 1B (Second-Moment Bound in Ambient Space (Strong)). *There exist scalars $\sigma_L, \sigma_b > 0$ such that for any unit-norm vector $u \in \mathbb{X}$ and any basis vector in $\{\phi_j\}_{j=1}^D$ we have the bounds*

$$\mathbb{E}\langle \phi_j, (L_i - L)u\rangle^2 \le \sigma_L^2 \|u\|^2, \quad \text{and} \tag{12a}$$

$$\mathbb{E}\langle \phi_j, b_i - b\rangle^2 \le \sigma_b^2. \tag{12b}$$

In other words, Assumption 1A guarantees that the random variable obtained by projecting the "noise" onto any of the basis vectors $\phi_1, \ldots, \phi_d$ in the subspace $\mathbb{S}$ has a bounded second moment. Assumption 1B further requires the projected noise onto any basis vector of the entire space $\mathbb{X}$ to have a bounded second moment. In Section 4, we show that there are various settings—including Galerkin methods and temporal difference methods—for which at least one of these assumptions is satisfied.

2.2. Examples

We now present some concrete examples to illustrate our general formulation. In particular, we discuss the problems of linear regression, temporal difference learning methods from reinforcement learning,[2] and Galerkin methods for solving partial differential equations.

2.2.1. Linear Regression on a Low-Dimensional Subspace. Our first example is the linear regression model when true parameter is known to lie approximately in a low-dimensional subspace. This example, although rather simple, provides a useful pedagogical starting point for the others to follow.

For this example, the underlying Hilbert space $\mathbb{X}$ from our general formulation is simply the Euclidean space $\mathbb{R}^D$, equipped with the standard inner product $\langle \cdot, \cdot \rangle$. We consider zero-mean covariates $X \in \mathbb{R}^D$ and a response $Y \in \mathbb{R}$, and our goal is to estimate the best-fitting linear model $x \mapsto \langle v, x\rangle$. In particular, the mean-square optimal fit is given by $v^* := \arg\min_{v \in \mathbb{R}^D} \mathbb{E}(Y - \langle v, X\rangle)^2$. From standard results on linear regression, this vector must satisfy the normal equations $\mathbb{E}[XX^\top]v^* = \mathbb{E}[YX]$. We assume that the second-moment matrix $\mathbb{E}[XX^\top]$ is nonsingular so that v^* is unique.

Let us rewrite the normal equations in a form consistent with our problem formulation. An equivalent definition of v^* is in terms of the fixed-point relation

$$v^* = \left(I - \frac{1}{\beta}\mathbb{E}[XX^\top]\right)v^* + \frac{1}{\beta}\mathbb{E}[YX], \tag{13}$$

where $\beta := \lambda_{\max}(\mathbb{E}[XX^\top])$ is the maximum eigenvalue. This fixed-point condition is a special case of our general equation (1) with the operator $L = I - \beta^{-1}\mathbb{E}[XX^\top]$ and vector $b = \beta^{-1}\mathbb{E}[YX]$. Note that we have

$$|||L|||_{\text{op}} = \left|\left|\left|I - \frac{1}{\beta}\mathbb{E}[XX^\top]\right|\right|\right|_{\text{op}} \leq 1 - \frac{\mu}{\beta} < 1,$$

where $\mu = \lambda_{\min}(\mathbb{E}[XX^\top]) > 0$ is the minimum eigenvalue of the covariance matrix.

In the well-specified setting of linear regression, we observe i.i.d. pairs $(X_i, Y_i) \in \mathbb{R}^D \times \mathbb{R}$ that are linked by the standard linear model

$$Y_i = \langle v^*, X_i\rangle + \varepsilon_i \quad \text{for } i = 1, 2, \ldots, n, \tag{14}$$

where ε_i denotes zero-mean noise with a finite second moment. Each such observation can be used to form the matrix-vector pair

$$L_i = I - \beta^{-1}X_iX_i^\top, \quad \text{and} \quad b_i = \beta^{-1}X_iY_i,$$

which is in the form of our assumed observation model.

Thus far, we have simply reformulated linear regression as a fixed-point problem. In order to bring in the projected aspect of the problem, let us suppose that the ambient dimension D is much larger than the sample size n but that we have the prior knowledge that v^* lies (approximately) within a known subspace $\mathbb{S}$ of $\mathbb{R}^D$-say, of dimension $d \ll D$. Our goal is then to approximate the solution to the associated projected fixed-point equation.

Using $\{\phi_j\}_{j=1}^d$ to denote an orthonormal basis of $\mathbb{S}$, the population-level projected linear equation (2) in this case takes the form

$$\mathbb{E}[(\Pi_{\mathbb{S}}X)(\Pi_{\mathbb{S}}X)^\top]\overline{v} = \mathbb{E}[Y \cdot \Pi_{\mathbb{S}}X]. \tag{15}$$

Thus, the population-level projected problem (15) corresponds to performing linear regression using the projected version of the covariates, thereby obtaining a vector of weights $\overline{v} \in \mathbb{S}$ in this low-dimensional space.

2.2.2. Galerkin Methods for Second-Order Elliptic Equations. We now turn to the Galerkin method for solving differential equations, a technique briefly described in Section 1. The general problem is to compute an approximate solution to a partial differential equation based on a limited number of noisy observations for the coefficients. Stochastic inverse problems of this type arise in various scientific and engineering applications (Arridge et al. [1], Nickl [44]).

For concreteness, we consider a second-order elliptic equation with Dirichlet boundary conditions.[3] Given a bounded, connected, and open set $\Omega \subseteq \mathbb{R}^m$ with unit Lebesgue measure, let $\partial\Omega$ denote its boundary. Consider the Hilbert space of functions

$$\mathbb{X} := \dot{\mathbb{H}}^1(\Omega) = \left\{v : \Omega \to \mathbb{R}, \ \int_\Omega \|\nabla v(x)\|_2^2 dx < \infty, \ v|_{\partial\Omega} = 0\right\},$$

equipped with the inner product $\langle u, v\rangle_{\dot{\mathbb{H}}^1} := \int_\Omega \nabla u(x)^\top \nabla v(x) dx$.

Given a symmetric matrix-valued function a and a square-integrable function $f \in \mathbb{L}^2$, the *boundary-value problem* is to find a function $v : \Omega \to \mathbb{R}$ such that

$$\begin{cases} \nabla \cdot (a(x)\nabla v(x)) + f = 0 & \text{in } \Omega, \\ v(x) = 0 & \text{on } \partial\Omega. \end{cases} \tag{16}$$

We impose a form of uniform ellipticity by requiring that $\mu I_m \preceq a(x) \preceq \beta I_m$ for some positive scalars $\mu \leq \beta$, valid uniformly over x.

The problem can be equivalently stated in terms of the elliptic operator $A := -\nabla \cdot (a\nabla)$; as shown in Appendix H.3.1 in the online supplementary file, the pair (A, f) *induces* a bounded, self-adjoint linear operator $\tilde{A}$ on $\mathbb{X}$ and a function $g \in \mathbb{X}$ such that the solution to the boundary value problem can be written as

$$v^* = \left(I - \frac{1}{\beta}\tilde{A}\right)v^* + \beta^{-1}g. \tag{17}$$

By construction, this is now an instance of our general fixed-point equation (1) with $L := I - \beta^{-1}\tilde{A}$ and $b := \beta^{-1}g$. Furthermore, our assumptions imply that $|||L|||_{\mathbb{X}} \le 1 - \mu/\beta$.

We consider a stochastic observation model that is standard in the literature (see, e.g., the paper by Giordano and Nickl [24]). Independently for each $i \in [n]$, let W_i denote an $m \times m$ symmetric random matrix with entries on the diagonal and upper diagonal given by i.i.d. standard Gaussian random variables. Let $w_i' \sim \mathcal{N}(0,1)$ denote a standard Gaussian random variable. Suppose now that we observe the pair $x_i, y_i \sim \mathcal{U}(\Omega)$; the observed values for the ith sample are then given by

$$(a_i, f_i) := (a(x_i) + W_i, f(y_i) + w_i') \quad \text{with} \quad x_i, y_i \sim \mathcal{U}(\Omega). \tag{18}$$

The unbiased observations (L_i, b_i) can then be constructed by replacing (a, f) with $(a_i\delta_{x_i}, f_i\delta_{y_i})$ in the above-mentioned constructions.

For such problems, the finite-dimensional projection not only serves as a fast and cheap way to compute solutions from simulation (Lung et al. [37]) but also makes the solution stable and robust to noise (Kaltenbacher et al. [25]). Given a finite-dimensional linear subspace $\mathbb{S} \subseteq \mathbb{X}$ spanned by orthogonal basis functions $(\phi_i)_{i=1}^d$, we consider the projected version of Equation (17), with the solution denoted by $\overline{v}$:

$$\overline{v} = \Pi_{\mathbb{S}}(L\overline{v} + b). \tag{19}$$

Straightforward calculation in conjunction with Lemma 9 shows that Equation (19) is equivalent to the conditions $\overline{v} \in \mathbb{S}$, and

$$\langle \tilde{A}\overline{v}, \phi_j \rangle_{\dot{\mathbb{H}}^1} = \langle g, \phi_j \rangle_{\dot{\mathbb{H}}^1} \quad \text{for all } j \in [d], \tag{20}$$

with the latter equality better known in the literature as the *Galerkin orthogonality condition* (Brenner and Scott [13]).

2.2.3. Temporal Difference Methods for Policy Evaluation. Our final example involves the policy evaluation problem in reinforcement learning. This is a special case of an instrumental variable method, as briefly introduced in Section 1. We require some additional terminology to describe the problem of policy evaluation. Consider a Markov chain on a state space $\mathcal{S}$ and a transition kernel $P : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$. It becomes a discounted Markov reward process when we introduce a reward function $r : \mathcal{S} \to \mathbb{R}$ and discount factor $\gamma \in (0,1)$. The goal of the policy evaluation problem is to estimate the value function, which is the expected, long-term, discounted reward accrued by running the process. The value function exists under mild assumptions such as boundedness of the reward and is given by the solution to the Bellman equation $v^* = \gamma P v^* + r$, which is a fixed-point equation of the form (1) with $L = \gamma P$ and $b = r$.

Throughout our discussion, we assume that the transition kernel P is ergodic and aperiodic so that its stationary distribution ξ is unique. We define $\mathbb{X}$ to be the Hilbert space $\mathbb{L}^2(\mathcal{S}, \xi)$, and for any pair of vectors $v, v' \in \mathbb{X}$, we define the inner product as follows:

$$\langle v, v' \rangle := \int_{\mathcal{S}} v(s)v'(s)d\xi(s).$$

In the special case of a finite state space, the Hilbert space $\mathbb{X}$ is a finite-dimensional Euclidean space with dimension $D = |\mathcal{S}|$ and equipped with a weighted ℓ_2-norm.

We consider the i.i.d. observation model in this paper. For each $i = 1, 2, \ldots, n$, suppose that we observe an independent tuple $(s_i, s_i^+, R_i(s_i))$ such that

$$s_i \sim \xi, \; s_i^+ \sim P(s_i, \cdot) \text{ and } \mathbb{E}[R_i(s_i)|s_i] = r(s_i). \tag{21}$$

The ith observation (L_i, b_i) is then obtained by plugging in these two observations to compute unbiased estimates of P and r, respectively.

A common practice in reinforcement learning is to employ *function approximation*, which in its simplest form involves solving a projected linear equation on a subspace. In particular, consider a set $\{\psi_1, \psi_2, \ldots, \psi_d\}$ of basis functions in $\mathbb{X}$, and suppose that they are linearly independent on the support of ξ. We are interested in projections onto the subspace $\mathbb{S} = \text{span}(\psi_1, \ldots, \psi_d)$ and in solving the population-level projected fixed-point equation (2), which takes the form

$$\overline{v} = \Pi_{\mathbb{S}}(\gamma P\overline{v} + r). \tag{22}$$

The basis functions ψ_i are not necessarily orthogonal, and it is common for the projection operation to be carried out in a somewhat nonstandard fashion. In order to describe this, it is convenient to write Equation (22) in the

projected space. For each $s \in \mathcal{S}$, let $\psi(s) = [\psi_1(s)\, \psi_2(s) \ldots \psi_d(s)]$ denote a vector in $\mathbb{R}^d$, and note that we may write $\overline{v}(s) = \psi(s)^\top \overline{\vartheta}$ for a vector of coefficients $\overline{\vartheta} \in \mathbb{R}^d$. Now observe that Equation (22) can be equivalently written in terms of the coefficient vector $\overline{\vartheta}$ as

$$\mathbb{E}_{s\sim\xi}[\psi(s)\psi(s)^\top]\overline{\vartheta} = \gamma \mathbb{E}_{s\sim\xi}[\mathbb{E}_{s^+\sim P(s,\cdot)}[\psi(s)\psi(s^+)^\top]]\overline{\vartheta} + \mathbb{E}_{s\sim\xi}[r(s)\psi(s)]. \tag{23}$$

Equation (23) is the population relation underlying the canonical *least squares temporal difference* learning method (Boyan [11], Bradtke and Barto [12]).

3. Main Results for General Projected Linear Equations

Having set up the problem and illustrating it with some examples, we now turn to the statements of our main results. We begin in Section 3.1 by stating an upper bound on the mean-squared error of a stochastic approximation scheme that uses Polyak–Ruppert averaging. We then discuss the form of this upper bound for various classes of operator L, with a specific focus on producing transparent bounds on the approximation factor. Section 3.2 is devoted to information-theoretic lower bounds that establish the sharpness of our upper bound.

3.1. Upper Bounds

In this section, we describe a standard stochastic approximation scheme for the problem based on combining ordinary stochastic updates with Polyak–Ruppert averaging (Polyak [46], Polyak and Juditsky [47], Ruppert [54]). In particular, given an oracle that provides observations (L_i, b_i), consider the stochastic recursion parameterized by a positive stepsize η:

$$v_{t+1} = (1-\eta)v_t + \eta\Pi_{\mathbb{S}}(L_{t+1}v_t + b_{t+1}), \quad \text{for } t = 1, 2, \ldots. \tag{24a}$$

This is a standard stochastic approximation scheme for attempting to solve the projected fixed-point relation. In order to improve it, we use the standard device of applying Polyak–Ruppert averaging so as to obtain our final estimate. For a given sample size $n \geq 2$, our final estimate $\hat{v}_n$ is given by taking the average of these iterates from time n_0 to n—that is,

$$\hat{v}_n := \frac{1}{n-n_0}\sum_{t=n_0+1}^{n} v_t. \tag{24b}$$

Here, the "burn-in" time n_0 is an integer parameter to be specified.

The stochastic approximation procedure (24) is defined in the entire space $\mathbb{X}$; note that it can be equivalently written as iterates in the projected space $\mathbb{R}^d$ via the recursion

$$\vartheta_{t+1} = (1-\eta)\vartheta_t + \eta(\Phi_d L_{t+1}\Phi_d^*\vartheta_t + \Phi_d b_{t+1}). \tag{25}$$

The original iterates can be recovered by applying the adjoint operator—that is, $v_t = \Phi_d^*\vartheta_t$ for $t = 1, 2, \ldots$.

3.1.1. A Finite-Sample Upper Bound. Having introduced the algorithm itself, we are now ready to provide a guarantee on its error. Two matrices play a key role in the statement of our upper bound. The first is the d-dimensional matrix $M := \Phi_d L \Phi_d^*$ that we introduced in Section 2.1. We show that the mean-squared error is upper bounded by the approximation error $\inf_{v\in\mathbb{S}}\|v - v^*\|^2$ along with a prefactor of the form

$$\alpha(M, s) = 1 + \lambda_{\max}((I-M)^{-1}(s^2 I_d - MM^T)(I-M)^{-T}) \tag{26}$$

for $s = |||L|||_{\text{op}}$. Our bounds also involve the quantity $\kappa(M) = \lambda_{\max}(M + M^T)/2$, which we abbreviate as κ when the underlying matrix M is clear from the context.

The second matrix is a covariance matrix, capturing the noise structure of our observations, given by

$$\Sigma^* := \text{cov}(\Phi_d(b_1 - b) + \Phi_d(L_1 - L)\overline{v}).$$

This matrix, along with the constants (σ_L, σ_b) from Assumption 1A, arises in the definition of two additional error terms:

$$\mathcal{E}_n(M, \Sigma^*) := \frac{\text{trace}((I-M)^{-1}\Sigma^*(I-M)^{-\top})}{n}, \quad \text{and} \tag{27a}$$

$$\mathcal{H}_n(\sigma_L, \sigma_b, \overline{v}) := \frac{\sigma_L}{(1-\kappa)^3}\left(\frac{d}{n}\right)^{3/2}(\|\overline{v}\|^2\sigma_L^2 + \sigma_b^2). \tag{27b}$$

As suggested by our notation, the error $\mathcal{H}_n(\sigma_L, \sigma_b, \overline{v})$ is a higher-order term, decaying as $n^{-3/2}$ in the sample size, whereas the quantity $\mathcal{E}_n(M, \Sigma^*)$ is the dominant source of statistical error. With this notation, we have the following.

Theorem 1. *Suppose that we are given n i.i.d. observations* $\{(L_i, b_i)\}_{i=1}^n$ *that satisfy the noise conditions in Assumption 1A. Then there are universal constants* (c_0, c) *such that for any sample size* $n \geq \frac{c_0 \sigma_L^2 d}{(1-\kappa)^2} \log^2\left(\frac{\|v_0 - \overline{v}\|^2 d}{1-\kappa}\right)$*, running the algorithm (24) with*

$$\text{step size } \eta = \frac{1}{c_0 \sigma_L \sqrt{dn}} \quad \text{and burn-in period } n_0 = n/2$$

yields an estimate $\hat{v}_n$ *such that*

$$\mathbb{E}\|\hat{v}_n - v^*\|^2 \leq (1+\omega) \cdot \alpha(M, |||L|||_{\mathbb{X}}) \inf_{v \in \mathbb{S}} \|v - v^*\|^2 + c\left(1 + \frac{1}{\omega}\right) \cdot \{\mathcal{E}_n(M, \Sigma^*) + \mathcal{H}_n(\sigma_L, \sigma_b, \overline{v})\}, \tag{28}$$

valid for any $\omega > 0$.

We prove this theorem in Section 5.1.

A few comments are in order. First, the quantity $\alpha(M, |||L|||_{\mathbb{X}}) \inf_{v \in \mathbb{S}} \|v - v^*\|^2$ is an upper bound on the approximation error $\|\overline{v} - v^*\|^2$ incurred by the (deterministic) projected fixed point $\overline{v}$. The prefactor $\alpha(M, |||L|||_{\mathbb{X}}) \geq 1$ measures the instance-specific deficiency of $\overline{v}$ relative to an optimal approximating vector from the subspace, and we provide a more in-depth discussion of this factor in Section 3.1.2 to follow. Note that Theorem 1 actually provides a family of bounds, indexed by the free parameter $\omega > 0$. By choosing ω arbitrarily close to 0, we can make the prefactor in front of $\inf_{v \in \mathbb{S}} \|v - v^*\|^2$ arbitrarily close to $\alpha(M, |||L|||_{\mathbb{X}})$—albeit at the expense of inflating the remaining error terms. In Theorem 2 to follow, we prove that the quantity $\alpha(M, |||L|||_{\mathbb{X}})$ is, in fact, the smallest approximation factor that can be obtained in any such bound.

The latter two terms in the bound (28) correspond to estimation error that arises from estimating $\overline{v}$ based on a set of n stochastic observations. Although there are two terms here in principle, we show in Corollary 1 to follow that the estimation error is dominated by the term $\mathcal{E}_n(M, \Sigma^*)$ under some natural assumptions. Note that the leading term $\mathcal{E}_n(M, \Sigma^*)$ scales with the local complexity for estimating $\overline{v}$, and we show in Theorem 3 that this term is also information-theoretically optimal. In Appendix I.2 in the online supplementary file, we perform additional simulation studies on the statistical error terms, showing that the actual performance of Polyak–Ruppert averaging estimator is accurately characterized by the instance-dependent analysis.

In the next subsection, we undertake a more in-depth exploration of the approximation factor in this problem, discussing prior work in the context of the term $\alpha(M, |||L|||_{\mathbb{X}})$ appearing in Theorem 1.

3.1.2. Detailed Discussion of the Approximation Error.

As mentioned in the introduction, upper bounds on the approximation factor have received significant attention in the literature, and it is interesting to compare our bounds.

3.1.2.1. Past Results.

In the case where $\gamma_{\max} := |||L|||_{\mathbb{X}} < 1$, the approximation-factor bound (6) was established by Tsitsiklis and Van Roy [60], via the following argument. Letting $\tilde{v} := \Pi_{\mathbb{S}}(Lv^* + b)$, we have

$$\begin{aligned}
\|\overline{v} - v^*\|^2 &\overset{\text{(i)}}{=} \|\overline{v} - \tilde{v}\|^2 + \|\tilde{v} - v^*\|^2 = \|\Pi_{\mathbb{S}}(L\overline{v} + b) - \Pi_{\mathbb{S}}(Lv^* + b)\|^2 + \|\tilde{v} - v^*\|^2 \\
&\overset{\text{(ii)}}{\leq} \|L\overline{v} - Lv^*\|^2 + \|\tilde{v} - v^*\|^2 \\
&\overset{\text{(iii)}}{\leq} \gamma_{\max}^2 \|\overline{v} - v^*\|^2 + \|\tilde{v} - v^*\|^2.
\end{aligned} \tag{29}$$

Step (i) uses the Pythagorean theorem, step (ii) follows from the nonexpansiveness of the projection operator, and step (iii) makes use of the contraction property of the operator L. Note that, by definition, we have $\alpha(M, |||L|||_{\mathbb{X}}) \leq (1 - |||L|||_{\mathbb{X}})^{-2}$, and so the approximation factor in Theorem 1 recovers the bound (6) in the worst case. In general, however, the factor $\alpha(M, |||L|||_{\mathbb{X}})$ can be significantly smaller. See Lemmas 1 and 2 to follow.

Yu and Bertsekas [71] derived two finer-grained upper bounds on the approximation factor; in terms of our notation, their bounds take the form

$$\begin{aligned}
\alpha_{\mathsf{YB}}^{(1)} &:= 1 + |||L|||_{\mathbb{X}}^2 \cdot \lambda_{\max}((I - M)^{-1}(I - M)^{-\top}), \\
\alpha_{\mathsf{YB}}^{(2)} &:= 1 + |||(I - \Pi_{\mathbb{S}} L)^{-1} \Pi_{\mathbb{S}} L \Pi_{\mathbb{S}^\perp}|||_{\mathbb{X}}^2.
\end{aligned}$$

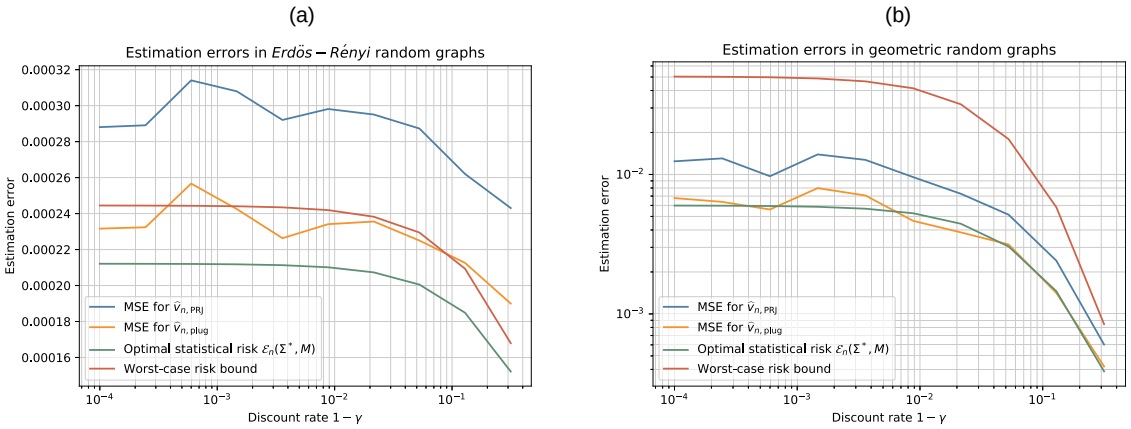
It is clear from the definition that $\alpha(M, |||L|||_{\mathbb{X}}) \le \alpha_{\mathsf{YB}}^{(1)}$, but $\alpha(M, |||L|||_{\mathbb{X}})$ can often provide an improved bound. This improvement is indeed significant, as will be shown shortly in Lemma 1. On the other hand, the term $\alpha_{\mathsf{YB}}^{(2)}$ is never larger than $\alpha(M, |||L|||_{\mathbb{X}})$, and it is indeed the smallest possible bound that depends only on L and *not* b. However, as pointed out by Yu and Bertsekas, the value of $\alpha_{\mathsf{YB}}^{(2)}$ is not easily accessible in practice, because it depends on the precise behavior of the operator L over the orthogonal complement $\mathbb{S}^{\perp}$. Thus, estimating the quantity $\alpha_{\mathsf{YB}}^{(2)}$ requires $O(D)$ samples. By contrast, the term $\alpha(M, |||L|||_{\mathbb{X}})$ depends only on the projected operator M and the operator norm $|||L|||_{\mathbb{X}}$. The former can be easily estimated using d samples and at smaller computational cost, whereas the latter is usually known a priori. The discussion in Section 4 fleshes out these distinctions.

3.1.2.2. A Simulation Study. In order to compare different upper bounds on the approximation factor, we conducted a simple simulation study on the problem of value function estimation, as previously introduced in Section 2.2.3. For this problem, the approximation factor $\alpha(M, \gamma)$ is computed more explicitly in Corollary 5. The Markov transition kernel is given by the simple random walk on a graph. We consider Gaussian random feature vectors and associate them with two different random graph models, Erdős–Rényi graphs and random geometric graphs. The details for these models are described and discussed in Appendix I.1 in the online supplementary file.

In Figure 1, we show the simulation results for the values of the approximation factor. Given a sample from preceding graphs and feature vectors, we plot the value of $\alpha(M, \gamma)$, $\alpha_{\mathsf{YB}}^{(1)}$ and $\alpha_{\mathsf{YB}}^{(2)}$ against the discount rate $1 - \gamma$, which ranges from 10^{-5} to $10^{-0.5}$. Note that the two plots use different scales: panel (a) is a linear-log plot, whereas panel (b) is a log-log plot. Figure 1 shows that the approximation factor $\alpha(M, \gamma)$ derived in Theorem 1 is always between $\alpha_{\mathsf{YB}}^{(1)}$ and $\alpha_{\mathsf{YB}}^{(2)}$. As mentioned before, the latter quantity depends on the particular behavior of the linear operator L in the subspace $\mathbb{S}^{\perp}$, which can be difficult to estimate. The improvement over $\alpha_{\mathsf{YB}}^{(1)}$, on the other hand, can be significant.

In the Erdős–Rényi model, all three quantities are bounded by a relatively small constant, regardless of the value of γ. The bound $\alpha(M, \gamma)$ is roughly at the midpoint between the bounds $\alpha_{\mathsf{YB}}^{(1)}$ and $\alpha_{\mathsf{YB}}^{(2)}$. On the other hand, the differences are much starker in the random geometric graph case: the bound improves over $\alpha_{\mathsf{YB}}^{(1)}$ by several orders of magnitude while being off from $\alpha_{\mathsf{YB}}^{(2)}$ by a factor of 10 for large γ. As we discuss shortly in Lemma 1, this is because the approximation factor $\alpha(M, \gamma)$ scales as $O(1 - \kappa(M))^{-1}$, whereas $\alpha_{\mathsf{YB}}^{(1)}$ scales as $O(1 - \kappa(M))^{-2}$, making a big difference in the case where the constant $\kappa(M)$ is large.

Figure 1. (Color online) Plots of various approximation factor as a function of the discount factor γ in the policy evaluation problem. (See Section 3.1.2.2 for a discussion.) (a) Results for an Erdős–Rényi random graph model with $n = 3{,}000$, projected dimension $d = 1{,}000$, and $a = 3$. The resulting number of vertices in the graph $\tilde{G}$ is 2,813. The value of $1 - \gamma$ is plotted in log-scale, and the value of approximation factor is plotted on the standard scale. (b) Results for a random geometric graph model with $n = 3{,}000$, projected dimension $d = 2$, and $r = 0.1$. The resulting number of vertices in the graph $\tilde{G}$ is 2,338. Both the discount rate $1 - \gamma$ and the approximation factor are plotted on the log-scale.



3.1.2.3. Some Useful Bounds on $\alpha(M, |||L|||_{\mathbb{X}})$. We conclude our discussion of the approximation factor with some bounds that can be derived under different assumptions on the operator L and its projected version M. The following lemma is useful in understanding the behavior of the approximation factor as a function of the contractivity properties of the operator L; this is particularly useful in analyzing convergence rates in numerical PDEs.

Lemma 1. *Consider a projected matrix $M \in \mathbb{R}^{d\times d}$ such that $(I - M)$ is invertible and $\kappa(M) < 1$.*
(a) *For any $s > 0$, we have the bound*

$$\alpha(M, s) \le 1 + |||(I - M)^{-1}|||_{\text{op}}^2 \cdot s^2 \le 1 + \frac{s^2}{(1 - \kappa(M))^2}. \tag{30a}$$

(b) *For $s \in [0, 1]$, we have*

$$\alpha(M, s) \le 1 + 2|||(I - M)^{-1}|||_{\text{op}} \le 1 + \frac{2}{1 - \kappa(M)}. \tag{30b}$$

See Appendix G.1 in the online supplementary file for the proof of this lemma.

A second special case, also useful, is when the matrix M is symmetric, a setting that appears in least squares regression, value function estimation in reversible Markov chains, and self-adjoint elliptic operators. The optimal approximation factor $\alpha(M, \gamma_{\max})$ can be explicitly computed in such cases.

Lemma 2. *Suppose that M is symmetric with eigenvalues $\{\lambda_j(M)\}_{j=1}^d$ such that $\lambda_{\max}(M) < 1$. Then for any $s > 0$, we have*

$$\alpha(M, s) = 1 + \max_{j=1,\dots,d} \frac{s^2 - \lambda_j^2}{(1 - \lambda_j)^2}. \tag{31}$$

See Appendix G.2 in the online supplementary file for the proof of this lemma.

Lemma 1 reveals that there is a qualitative shift between the nonexpansive case $|||L|||_{\mathbb{X}} \le 1$ and the complementary expansive case. In the latter case, the optimal approximation factor always scales as $O(1 - \kappa(M))^{-2}$, but below the threshold $|||L|||_{\mathbb{X}} = 1$, the approximation factor drastically improves to become $O(1 - \kappa(M))^{-1}$. It is worth noting that both bounds can be achieved up to universal constant factors. In the context of differential equations, the bound of the form (*a*) in Lemma 1 is known as Céa's lemma (Céa [16]), which plays a central role in the convergence rate analysis of the Galerkin methods for numerical differential equations. However, the instance-dependent approximation factor $\alpha(M, |||L|||_{\mathbb{X}})$ can often be much smaller: the global coercive parameter needed in Céa's estimate is replaced by the bounds on the behavior of the operator L in the finite-dimensional subspace. Part (b) in Lemma 1 generalizes Céa's energy estimate from the symmetric positive-definite case to the general nonexpansive setting. See Corollary 4 for a more detailed discussion on the consequences of our results to elliptic PDEs.

Lemmas 1 and 2 yield the following corollary of the general bound (28) under different conditions on the operator L.

Corollary 1. *Under the conditions of Theorem 1 and given a sample size $n \ge \frac{c_0 \sigma_L^2 d}{(1-\kappa)^2} \log^2\left(\frac{\|v_0 - \overline{v}\|^2 d}{1-\kappa}\right)$, we have the following.*
(a) *There is a universal positive constant c such that*

$$\mathbb{E}\|\hat{v}_n - v^*\|^2 \le c\left\{\frac{|||L|||_{\mathbb{X}}^2}{(1 - \kappa(M))^2} \cdot \inf_{v\in\mathbb{S}} \|v - v^*\|^2 + \frac{(\sigma_b^2 + \sigma_L\|\overline{v}\|^2)}{(1 - \kappa(M))^2} \frac{d}{n}\right\} \tag{32a}$$

for any operator L and its associated projected operator $M = \Phi_d L \Phi_d^*$.
(b) *Moreover, when L is nonexpansive ($|||L|||_{\mathbb{X}} \le 1$), we have*

$$\mathbb{E}\|\hat{v}_n - v^*\|^2 \le c\left\{\frac{1}{1 - \kappa(M)} \cdot \inf_{v\in\mathbb{S}} \|v - v^*\|^2 + \frac{(\sigma_b^2 + \sigma_L\|\overline{v}\|^2)}{(1 - \kappa(M))^2} \frac{d}{n}\right\}. \tag{32b}$$

See Appendix C in the online supplementary file for the proof of this claim.

As alluded to before, the simplified form of Corollary 1 no longer has an explicit higher-order term, and the statistical error now scales at the parametric rate d/n. It is worth noting that the lower bound on n required in the assumption of the corollary is a mild requirement: in the absence of such a condition, the statistical error term $\frac{(\sigma_b^2 + \sigma_L\|\overline{v}\|^2)}{(1-\kappa)^2} \frac{d}{n}$ in both bounds would blow up, rendering the guarantee vacuous.

3.2. Lower Bounds

In this section, we establish information-theoretic lower bounds on the approximation factor, as well as the statistical error. Our eventual result (in Corollary 2) shows that the first two terms appearing in Theorem 1 are both optimal in a certain instance-dependent sense. However, a precise definition of the local neighborhood of instances over which the lower bound holds requires some definitions. In order to motivate these definitions more transparently and naturally arrive at both terms of the bound, the following section presents individual bounds on the approximation and estimation errors and then combines them to obtain Corollary 2.

3.2.1. Lower Bounds on the Approximation Error. As alluded to in the preceding, the first step involved in a lower bound is a precise definition of the collection of problem instances over which it holds; let us specify a natural such collection for lower bounds on the approximation error. Each problem instance is specified by the joint distribution of the observations (L_i, b_i), which implicitly specifies a pair of means $(L, b) = (\mathbb{E}[L_i], \mathbb{E}[b_i])$. For notational convenience, we define this class by first defining a collection comprising instances specified solely by the mean pair (L, b) and then providing restrictions on the distribution of (L_i, b_i). Let us define the first such component. For a given matrix $M_0 \in \mathbb{R}^{d\times d}$ and vector $h_0 \in \mathbb{R}^d$, write

$$\mathbb{C}_{\mathsf{approx}}(M_0, h_0, D, \delta, \gamma_{\max}) := \left\{ (L,b) \;\middle|\; \begin{array}{c} |||L|||_{\mathbb{X}} \le \gamma_{\max}, \quad \mathcal{A}(\mathbb{S}, v^*) \le \delta^2, \quad \dim(\mathbb{X}) = D, \\ \Phi_d L \Phi_d^* = M_0, \quad \text{and} \quad \Phi_d b = h_0. \end{array} \right\}.$$

In other words, this is a collection of all instances of the pair $(L, b) \in \mathfrak{L} \times \mathbb{R}^D$ whose projections onto the subspace of interest are fixed to be the pair (M_0, h_0) and whose approximation error is less than δ^2. In addition, the operator L satisfies a certain bound on its operator norm.

Having specified a class of (L, b) pairs, we now turn to the joint distribution over the pair of observations (L_i, b_i), which we denote for convenience by $\mathbb{P}_{L,b}$. Now define the collection of instances

$$\mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b) := \{\mathbb{P}_{L,b} \,|(L_i, b_i) \text{satisfies Assumption 1(B)with constants } (\sigma_L, \sigma_b)\}.$$

This is simply the class of all distributions such that our observations satisfy Assumption 1B with prespecified constants. As a point of clarification, it is useful to recall that our upper bound in Theorem 1 only needed Assumption 1A to hold, and we could have chosen to match this by defining the $\mathbf{G}_{\mathsf{var}}$ under Assumption 1A. We comment further on this issue following the theorem statement.

We are now ready to state Theorem 2, which is a lower bound on the worst-case approximation factor over all problem instances such that $(L, b) \in \mathbb{C}_{\mathsf{approx}}(M_0, h_0, D, \delta, \gamma_{\max})$ and $\mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)$. Note that such a collection of problem instances is indeed *local* around the pair (M_0, h_0). Two settings are considered in the statement of the theorem: *proper* estimators when $\hat{v}_n$ is restricted to take values in the subspace $\mathbb{S}$ and *improper* estimators where $\hat{v}_n$ can take values in the entire space $\mathbb{X}$. We use $\hat{\mathcal{V}}_{\mathbb{S}}$ and $\hat{\mathcal{V}}_{\mathbb{X}}$ to denote the class of proper and improper estimators, respectively. Finally, we use the shorthand $\mathbb{C}_{\mathsf{approx}} \equiv \mathbb{C}_{\mathsf{approx}}(M_0, h_0, D, \delta, \gamma_{\max})$ for convenience.

Theorem 2. *Suppose $M_0 \in \mathbb{R}^{d\times d}$ is a matrix such that $I - M_0$ is invertible and that the scalars (σ_L, σ_b) are such that $\sigma_L \ge \gamma_{\max}$ and $\sigma_b \ge \delta$. If the ambient dimension satisfies $D \ge d + \frac{12}{\omega} n^2$ for some scalar $\omega \in (0,1)$, then we have the lower bounds*

$$\inf_{\hat{v}_n \in \hat{\mathcal{V}}_{\mathbb{S}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\hat{v}_n - v^*\|^2 \ge (1-\omega) \cdot \alpha(M_0, \gamma_{\max}) \cdot \delta^2 \quad \textit{and} \tag{33a}$$

$$\inf_{\hat{v}_n \in \hat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\hat{v}_n - v^*\|^2 \ge (1-\omega) \cdot (\alpha(M_0, \gamma_{\max}) - 1) \cdot \delta^2. \tag{33b}$$

See Section 5.2 for the proof of this claim.

A few remarks are in order. First, Theorem 2 shows that the approximation factor upper bound in Theorem 1 is information-theoretically optimal in the instance-dependent sense: in the case of proper estimators, the upper and lower bounds can be made arbitrarily close by choosing the constant ω arbitrarily small in both theorems. Both bounds depend on the projected matrix M_0, characterizing the fundamental impact of the geometry in the projected space on the complexity of the estimation problem. The lower bound for improper estimators is slightly smaller, but for most practical applications, we have $\alpha(M_0, \gamma_{\max}) \gg 1$, and so this result should be viewed as almost equivalent.

Second, note that we may also extract a worst-case lower bound on the approximation factor from Theorem 2. Indeed, for a scalar $\gamma_{\max} \in (0,1)$, consider the family of instances in the aforementioned problem classes satisfying $|||L|||_{\mathbb{X}} \le \gamma_{\max}$. Setting $M_0 = \gamma_{\max}^2 I_d$ and applying Theorem 2, we see that (in a worst-case sense over this class) the

risk of any estimator is lower bounded by $\frac{1}{1-\gamma_{\max}^2}\mathcal{A}(\mathbb{S}, v^*)$. This establishes the optimality of the classical worst-case upper bound (6).

Third, notice that the theorem requires the noise variances (σ_L, σ_b) to be large enough, and this is a natural requirement in spite of the fact that we seek lower bounds on the approximation error. Indeed, in the extreme case of noiseless observations, we have access to the population pair (L, b) with a single sample and can compute both v^* and its projection onto the subspace $\mathbb{S}$ without error. From a more quantitative standpoint, it is worth noting that our requirements $\sigma_L \geq \gamma_{\max}$ and $\sigma_b \geq \delta$ are both mild, because the scalars $\gamma_{\max}$ and δ are typically order 1 quantities. Indeed, if both of these bounds held with equality, then Corollary 1 yields that the statistical error would be of the order $O(d/n)$ and so strictly smaller than the approximation error we hope to capture.[4]

Observe that Theorem 2 requires the ambient dimension D to be larger than n^2. As mentioned in the introduction, we should not expect any nontrivial approximation factor when $n \geq D$, but this leaves open the regime $n \ll D \ll n^2$. Is a smaller approximation factor achievable when D is not extremely large? We revisit this question in Section 3.2.4, showing that although there are some quantitative differences in the lower bound, the qualitative nature of the message remains unchanged.

Regarding our noise assumptions, it should first be noted that the class of instances satisfying Assumption 1A is strictly larger than the corresponding class satisfying Assumption 1B, and so our lower bound extends immediately to the former case. Second, it is important to note that imposing only Assumption 1A would, in principle, allow the noise in the orthogonal complement $\mathbb{S}^\perp$ to grow in an unbounded fashion, and one should expect that it is indeed optimal to return an estimate of the projected fixed point $\overline{v}$. As such, Theorem 2 constitutes a more meaningful lower bound, because we operate instead under the stronger Assumption 1B and enforce second-moment bounds on the noise not only for basis vectors in $\mathbb{S}$ but also for its orthogonal complement. Assumption 1B allows for other natural estimators: for instance, the plug-in estimator of v^* via the original fixed-point equation (1) would now incur finite error. Our lower bound—which operates under the stronger assumption and is thus more challenging to establish—shows that the stochastic approximation estimator analyzed in Theorem 1 is optimal *even if* the noise in $\mathbb{S}^\perp$ behaves as well as that in $\mathbb{S}$.

3.2.2. Lower Bounds on the Estimation Error. We now turn to establishing a minimax lower bound on the estimation error that matches the statistical error term in Theorem 1. This lower bound takes a slightly different form from Theorem 2: rather than studying the total error $\|\hat{v}_n - v^*\|$ directly, we establish a lower bound on the error $\|\hat{v}_n - \overline{v}\|$ instead.

Indeed, the latter term is more meaningful to study in order to characterize the estimation error—which depends on the sample size n—because for large sample sizes, the total error $\|\hat{v}_n - v^*\|$ will be dominated by a constant approximation error. As we demonstrate shortly, the term $\|\hat{v}_n - \overline{v}\|$ depends on noise covariance and the geometry of the matrix M_0 in the *projected space* while having the desired dependence on the sample size n. It is worth noting also that this automatically yields a lower bound on the error $\|\hat{v}_n - v^*\|$ when we have $\overline{v} = v^*$.

We are now ready to prove a local minimax lower bound for estimating $\overline{v} \in \mathbb{S}$, which is given by the solution to the projected linear equation $\overline{v} = \Pi_{\mathbb{S}}(L\overline{v} + b)$. Although our objective is to prove a local lower bound around each pair $(L_0, b_0) \in \mathfrak{L} \times \mathbb{X}$, the fact that we are estimating $\overline{v}$ implies that it suffices to define our set of local instances in the d-dimensional space of projections. In particular, our mean parameters (L, b) are specified by those pairs for which $\Phi_d L \Phi_d^*$ is close to $M_0 := \Phi_d L_0 \Phi_d^*$ and $\Phi_d b$ is close to $h_0 := \Phi_d b_0$. Specifically, let $\overline{v}_0$ denote the solution to the projected linear equation $\overline{v}_0 = \Pi_{\mathbb{S}}(L_0\overline{v}_0 + b_0)$, and define the neighborhood

$$\mathfrak{N}(M_0, h_0) := \left\{ (M', h') : |||M' - M_0|||_F \leq \sigma_L \sqrt{\frac{d}{n}} \text{ and } \|h' - h_0\|_2 \leq \sigma_b \sqrt{\frac{d}{n}} \right\}, \tag{34}$$

which, in turn, defines a local class of problem instances (L, b) given by

$$\mathbb{C}_{\text{est}} := \{(L, b) | (\Phi_d L \Phi_d^*, \Phi_d b) \in \mathfrak{N}(M_0, h_0)\}.$$

We have thus specified our local neighborhood in terms of the mean pair (L, b), and as before, it remains to define a local class of distributions on these instances. To this end, define the class

$$\mathbf{G}_{\text{cov}}(\Sigma_L, \Sigma_b, \sigma_L, \sigma_b) := \mathbf{G}_{\text{var}}(\sigma_L, \sigma_b) \cap \{\mathbb{P}_{L,b} \,|\, \text{cov}(\Phi_d(b_1 - b)) \preceq \Sigma_b \quad \text{and} \quad \text{cov}(\Phi_d(L_1 - L)\overline{v}_0) \preceq \Sigma_L\}, \tag{35}$$

which corresponds to distributions on the observation pair (L_i, b_i) that satisfy Assumption 1B and whose "effective noise" covariances are dominated by the positive semidefinite (PSD) matrices Σ_L and Σ_b.

Note that Assumption 1B implies the diagonal elements of preceding two covariance matrices are bounded by σ_b^2 and $\sigma_L^2\|\overline{v}_0\|^2$, respectively. In order to avoid conflicts between assumptions, we assume throughout that for all indices $j \in [d]$, the diagonal entries of the covariance matrices satisfy the conditions

$$(\Sigma_b)_{j,j} \le \sigma_b^2 \qquad \text{and} \qquad (\Sigma_L)_{j,j} \le \sigma_L^2\|\overline{v}_0\|^2. \tag{36}$$

We then have the following theorem for the estimation error $\|\hat{v}_n - \overline{v}\|$, where we use the shorthand $\mathbf{G}_{\mathsf{cov}} \equiv \mathbf{G}_{\mathsf{cov}}(\Sigma_L, \Sigma_b, \sigma_L, \sigma_b)$ for brevity.

Theorem 3. *Under the above-mentioned setup, suppose the matrix $I - M_0$ is invertible, and suppose that $n \ge 16\sigma_L^2 |||(I-M_0)^{-1}|||_{\mathrm{op}}^2 d$. Then there is a universal constant $c > 0$ such that*

$$\inf_{\hat{v}_n \in \hat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{est}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\hat{v}_n - \overline{v}\|^2 \ge c \cdot \mathcal{E}_n(M_0, \Sigma_L + \Sigma_b).$$

See Appendix D in the online supplementary file for the proof of this claim.

The estimation error lower bound in Theorem 3 is the worst-case instantiation of the statistical error term $\mathcal{E}_n(M, \Sigma^*)$ in Theorem 1 within the local problem class, up to a universal constant. Indeed, in the asymptotic limit $n \to \infty$, the regularity of the problem can be leveraged in conjunction with classical Le Cam theory (see, e.g., van der Vaart [63]) to show that the asymptotic optimal limiting distribution is a Gaussian law with covariance $(I-M)^{-1}\Sigma^*(I-M)^{-\top}$. (See the paper by Khamaru et al. [26] for a detailed analysis of this type in the special case of policy evaluation in tabular Markov decision processes.) This optimality result holds in a "local" sense: it is minimax optimal in a small neighborhood of radius $O(1/\sqrt{n})$ around a given problem instance (M_0, h_0). Theorem 3, on the other hand, is nonasymptotic, showing that a similar result holds provided n is lower bounded by an explicit, problem-dependent quantity of the order $\sigma_L^2 d |||(I-M_0)^{-1}|||_{\mathrm{op}}^2$. This accommodates a broader range of sample sizes than the upper bound in Theorem 1.

3.2.3. Combining the Bounds. Having presented separate lower bounds on the approximation and estimation errors in conjunction with definitions of local classes of instances over which they hold, we are now ready to present a corollary that combines the two lower bounds in Theorems 2 and 3.

We begin by defining the local classes of instances over which our combined bound holds. Given a matrix-vector pair (M_0, h_0), covariance matrices (Σ_L, Σ_b), ambient dimension $D > 0$, and scalars $\delta, \gamma_{\max}, \sigma_L, \sigma_b > 0$, we begin by specifying a collection of mean pairs (L, b) via

$$\mathbb{C}_{\mathsf{final}}(M_0, h_0, D, \delta, \gamma_{\max}) := \bigcup_{(M', h') \in \mathfrak{N}_n(M_0, h_0)} \mathbb{C}_{\mathsf{approx}}(M', h', D, \delta, \gamma_{\max}). \tag{37}$$

Clearly, this represents a natural combination of the classes $\mathbb{C}_{\mathsf{approx}}$ and $\mathbb{C}_{\mathsf{est}}$ introduced in the preceding. We use the shorthand $\mathbb{C}_{\mathsf{final}}$ for this class for brevity. Our collection of distributions $\mathbb{P}_{L,b}$ is still given by the class $\mathbf{G}_{\mathsf{cov}}$ from Equation (35).

With these definitions in hand, we are now ready to state our combined lower bound.

Corollary 2. *Under the above-mentioned setup, suppose that the pair (σ_L, σ_b) satisfies the conditions in Theorem 2 and* Equation (36) *and that the matrix M_0 satisfies $|||M_0|||_{\mathrm{op}} \le \gamma_{\max} - \sigma_L\sqrt{d/n}$. Moreover, suppose that the sample size and ambient dimension satisfy $n \ge 16\sigma_L^2 |||(I-M_0)^{-1}|||_{\mathrm{op}}^2 d$ and $D \ge d + 36n^2$, respectively. Then the following minimax lower bound holds for a universal positive constant c:*

$$\inf_{\hat{v}_n \in \hat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\mathsf{final}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{cov}}}} \mathbb{E}\|\hat{v}_n - v^*\|^2 \ge c \cdot \{(\alpha(M_0, \gamma_{\max}) - 1) \cdot \delta^2 + \mathcal{E}_n(M_0, \Sigma_L + \Sigma_b)\}.$$

We prove this corollary in Appendix E in the online supplementary file. It is a relatively straightforward consequence of combining Theorems 2 and 3.

The combined lower bound matches the expression $\alpha(M_0, \gamma_{\max})\mathcal{A}(\mathbb{S}, v^*) + \mathcal{E}_n(M_0, \Sigma_L + \Sigma_b)$, given by the first two terms of Theorem 1, up to universal constant factors. Recall from our discussion of Theorem 1 that the high-order term $\mathcal{H}_n(\sigma_L, \sigma_b, \overline{v})$ represents the "optimization error" of the stochastic approximation algorithm, which depends on the coercive condition $\kappa(M_0)$ instead of the natural geometry $I - M_0$ of the problem. Although we do not expect this term to appear in an information-theoretic lower bound, the leading estimation error term

$\mathcal{E}_n(M_0, \Sigma_L + \Sigma_b)$ will dominate the high-order term when the sample size n is large enough. For such a range of n, the bound in Theorem 1 is information-theoretically optimal in the local class specified in the preceding. More broadly, consider the class of all instances satisfying Assumption 1B, with $\kappa(M) \le \kappa$ and $|||L|||_{\mathbb{X}} \le 1$. Then the bound in Theorem 1 is optimal, in a worst-case sense, over this class as long as the sample size exceeds the threshold $\frac{c\sigma_L^2}{(1-\kappa)^2} d$.

3.2.4. The Intermediate Regime. It remains to tie up some loose ends. Note that the lower bound in Theorem 2 requires a condition $D \gg n^2$. On the other hand, it is easy to see that the approximation factor can be made arbitrarily close to 1 when $n \gg D$. (For example, one could run the estimator based on stochastic approximation and averaging—which was analyzed in Theorem 1—with the entire Euclidean space $\mathbb{X}$ and project the resulting estimate onto the subspace $\mathbb{S}$.) In the middle regime $n \ll D \ll n^2$, however, it is not clear which estimator is optimal.

In the following theorem, we present a lower bound for the approximation factor in this intermediate regime, which establishes the optimality of Theorem 1 up to a constant factor.

Theorem 4. *Suppose $M_0 \in \mathbb{R}^{d\times d}$ is a matrix such that $I - M_0$ is invertible and that the scalars (σ_L, σ_b) satisfy $\sigma_L \ge 1 + \gamma_{\max}$ and $\sigma_b \ge \delta$. If the ambient dimension satisfies $D \ge d + 3qn^{1+1/q}$ for some integer $q \in \left[2, \log n \wedge \frac{1}{\sqrt{2(1-\gamma_{\max}\wedge 1)}}\right]$, then we have the lower bound*

$$\inf_{\hat{v}_n \in \hat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\text{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\text{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\hat{v}_n - v^*\|^2 \ge \frac{\alpha(M, \gamma_{\max}) - 1}{4q^2} \cdot \delta^2.$$

See Appendix F in the online supplementary file for the proof of this theorem.

Theorem 4 resolves the gap in the intermediate regime, up to a constant factor that depends on q. In particular, the stochastic approximation estimator (24) for projected equations still yields a near-optimal approximation factor. Compared with Theorem 2, Theorem 4 weakens the requirement on the ambient dimension D and covers the entire regime $D \gg n$. Furthermore, using the same arguments as in Corollary 2, this theorem can also be combined with Theorem 3 to obtain the following lower bound in the regime $D \ge d + 3qn^{1+1/q}$ for any integer $q > 0$:

$$\inf_{\hat{v}_n \in \hat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\text{final}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\text{cov}}}} \mathbb{E}\|\hat{v}_n - v^*\|^2 \ge c \cdot \left\{ \frac{\alpha(M_0, \gamma_{\max}) - 1}{q^2} \cdot \delta^2 + \mathcal{E}_n(M_0, \Sigma_L + \Sigma_b) \right\}.$$

Let us summarize our approximation factor lower bounds in the various regimes. Consider a sequence of problem instances $(\mathbb{P}_{L,b}^{(n)})_{n=1}^{\infty}$ with increasing ambient dimension D_n. Let the projected dimension d, noise variances (σ_L, σ_b), oracle error δ, projected matrix $\Phi_d L^{(n)} \Phi_d^* = M$, and operator norm bound $|||L|||_{\mathbb{X}} \le \gamma_{\max}$ all be fixed. Table 1 presents a combination of our results from Theorems 1, 2, and 4; our results suggest that the optimal approximation factor exhibits a "slow" phase transition phenomenon. It is an interesting open question whether the phase transition is sharp and to identify the asymptotically optimal approximation factor in the regime $\lim_{n\to\infty} \frac{\log D_n}{\log n} = 1$ because our lower bounds do not apply in this linear regime.

4. Consequences for Specific Models

We now discuss the consequences of our main theorems for the three examples introduced in Section 2.2. For brevity, we state only upper bounds for the first two examples; our third example for temporal difference learning methods includes both upper and lower bounds.

Table 1. Bounds on the approximation factor $\frac{\mathbb{E}\|\hat{v}_n - v^*\|^2}{\mathcal{A}(\mathbb{S}, v^*)}$ for proper estimators in different ranges of ambient dimension. Here, $c_q \in (0,1)$ represents a constant depending only on the aspect ratio q.

$q = \lim_{n\to\infty} \frac{\log D_n}{\log n}$	$[2, \infty)$	$(1, 2)$	$(0, 1)$
Lower bound	$\alpha(M_0, \gamma_{\max})$	$c_q \cdot \alpha(M_0, \gamma_{\max})$	1
Upper bound	$\alpha(M_0, \gamma_{\max})$	$\alpha(M_0, \gamma_{\max})$	1

4.1. Linear Regression

Recall the setting of linear regression[5] from Section 2.2.1, including our i.i.d. observation model (14). We assume bounds on the second moment of ε and fourth moment of X—namely, the existence of some $\varsigma > 0$ such that

$$\mathbb{E}\langle u, X\rangle^4 \le \varsigma^4 \quad \text{and} \quad \mathbb{E}(\varepsilon^2) \le \varsigma^2 \qquad \text{for all } u \in \mathbb{S}^{D-1}. \tag{38}$$

These conditions ensure that Assumption 1A is satisfied with $(\sigma_L, \sigma_b) = (\beta^{-1}\varsigma^2, \beta^{-1}\varsigma^2)$. Recall that the (unprojected) covariance matrix satisfies the PSD relations $\mu I \preceq \mathbb{E}[XX^\top] \preceq \beta I$, and define the d-dimensional covariance matrix $\Sigma := \mathbb{E}[(\Phi_d X)(\Phi_d X)^\top]$ for convenience.

In this case, our stochastic approximation iterates (24a) take the form

$$v_{t+1} = v_t - \eta(\Pi_{\mathbb{S}} X_{t+1} X_{t+1}^\top \Pi_{\mathbb{S}} v_t + Y_{t+1}\Pi_{\mathbb{S}} X_{t+1}) \quad \text{for all } t = 0, 1, 2, \ldots, \tag{39}$$

and we take the averaged iterates $\hat{v}_n := \frac{2}{n}\sum_{t=n/2}^{n-1} v_t$. For this procedure, we have the following guarantee.

Corollary 3. *Suppose that we have n i.i.d. observations $\{(X_i, Y_i)\}_{i=1}^n$ from the model (14) satisfying the moment conditions (38). Then there are universal positive constants (c, c_0) such that given a sample size $n \ge \frac{c_0\varsigma^4 d}{\lambda_{\min}^2(\Sigma)}\log^2\left(\frac{\beta}{\mu}\|v_0 - \overline{v}\|_2^2 d\right)$, if the stochastic approximation scheme (39) is run with step size $\eta = \frac{1}{c_0\varsigma^2\sqrt{dn}}$, then the averaged iterate satisfies the bound*

$$\mathbb{E}\|\hat{v}_n - v^*\|_2^2 \le (1+\omega)\cdot\alpha\left(I_d - \frac{\Sigma}{\beta}, 1 - \frac{\mu}{\beta}\right)\mathcal{A}(\mathbb{S}, v^*) + c\cdot\frac{\operatorname{trace}(\Sigma^{-1})\cdot\mathbb{E}(\varepsilon^2)}{\omega n} + \frac{c}{\omega}\left(\frac{\varsigma^2}{\lambda_{\min}(\Sigma)}\cdot\sqrt{\frac{d}{n}}\right)^3$$

for any $\omega > 0$.

This result is a direct consequence of Theorem 1 in application to this model.

Note that the statistical error term $\frac{\operatorname{trace}(\Sigma^{-1})\cdot\mathbb{E}(\varepsilon^2)}{n}$ in this case corresponds to the classical statistical rates for linear regression in this low-dimensional subspace. The approximation factor, by Lemma 2, admits the closed-form expression

$$\alpha\left(I_d - \frac{\Sigma}{\beta}, 1 - \frac{\mu}{\beta}\right) = \max_{j\in[d]}\frac{\mu^2 + 2\beta(\lambda_j - \mu)}{\lambda_j^2},$$

where $\{\lambda_j\}_{j=1}^d$ denote the eigenvalues of the matrix Σ. Because $\lambda_j \in [\mu, \beta]$ for each $j \in [d]$, the approximation factor is at most of the order $O\left(\frac{\beta}{\lambda_{\min}(\Sigma)}\right)$.

Compared with known sharp oracle inequalities for linear regression (e.g., Rigollet and Hütter [51]), the approximation factor in our bound is not 1 but rather a problem-dependent quantity. This is because we study the *estimation error* under the standard Euclidean metric $\|\cdot\|_2$, as opposed to the *prediction error* under the data-dependent metric $\|\cdot\|_{L^2(P_X)}$. When the covariance matrix $\mathbb{E}[XX^\top]$ is identity, the approximation factor $\alpha\left(I_d - \frac{\Sigma}{\beta}, 1 - \frac{\mu}{\beta}\right)$ is equal to 1, recovering classical results. Another error metric of interest, motivated by applications such as transfer learning (Li et al. [35]), is the prediction error when the covariates X follow a different distribution Q. For such a problem, the aforementioned result can be modified straightforwardly by choosing the Hilbert space $\mathbb{X}$ to be $\mathbb{R}^D$, equipped with the inner product $\langle u, v\rangle := u^\top(\mathbb{E}_Q[XX^\top])^{-1}v$.

4.2. Galerkin Methods

We now return to the example of Galerkin methods, as previously introduced in Section 2.2.2, with the i.i.d. observation model (18). We assume the basis functions $\phi_1, \ldots, \phi_d$ to have a uniformly bounded function value and gradient, and we define the scalars

$$\sigma_L := \left(1 + \frac{2}{\beta}\right)\max_{j\in[d]}\sup_{x\in\Omega}\|\nabla\phi_j(x)\|_2 \quad \text{and} \quad \sigma_b := \frac{\|f\|_{\mathbb{L}^2} + 1}{\beta}\max_{j\in[d]}\sup_{x\in\Omega}|\phi_j(x)|. \tag{40}$$

These boundedness conditions are naturally satisfied by many interesting basis functions such as the Fourier basis[6] and ensure—we verify this concretely in the proof of Corollary 4 to follow—that our observation model satisfies Assumption 1A with parameters (σ_L, σ_b).

Taking the finite-dimensional representation $v = \vartheta^\top\phi$, the stochastic approximation estimator for solving Equation (19) is given by

$$\vartheta_{t+1} = \vartheta_t - \beta^{-1}\eta(\nabla\phi(x_{t+1})^\top a_{t+1}\nabla\phi(x_{t+1})\vartheta_t - f_{t+1}\phi(y_{t+1})) \quad \text{for } t = 0, 1, \ldots,$$

$$\hat{\vartheta}_n := \frac{2}{n}\sum_{t=n/2}^{n-1}\vartheta_t, \quad \text{and} \quad \hat{v}_n := \hat{\vartheta}_n^\top\phi.$$

In order to state our statistical guarantees for $\hat{v}_n$, we define the following matrices:

$$M := I_d - \beta^{-1} \int_\Omega \nabla\phi(x)^\top a(x) \nabla\phi(x) dx,$$

$$\Sigma_L := \frac{1}{\beta^2} \int_\Omega (\nabla\phi)^\top a \nabla\overline{v} (\nabla\overline{v})^\top a \nabla\phi \, dx - \frac{1}{\beta^2} \left(\int_\Omega (\nabla\phi)^\top a \nabla\overline{v} \, dx \right) \left(\int_\Omega (\nabla\phi)^\top a \nabla\overline{v} \, dx \right)^\top$$
$$+ \frac{1}{\beta^2} \int_\Omega (\nabla\phi)^\top \left[(\nabla\overline{v})(\nabla\overline{v})^\top + \mathrm{diag}(\|\nabla\overline{v}\|_2^2 - (\partial_j \overline{v})^2)_{j=1}^m \right] (\nabla\phi) dx,$$

$$\Sigma_b := \frac{1}{\beta^2} \int_\Omega (f(x)^2 + 1) \phi(x) \phi(x)^\top \, dx - \frac{1}{\beta^2} \left(\int_\Omega f(x) \phi(x) dx \right) \left(\int_\Omega f(x) \phi(x) dx \right)^\top.$$

With these definitions in hand, we are ready to state the consequence of our main theorems to the estimation problem of elliptic equations.

Corollary 4. *Under the above-mentioned setup, there are universal positive constants* (c, c_0) *such that if* $n \geq \frac{c_0 \sigma_L^2 d}{(1-\kappa(M))^2} \log^2\left(\frac{\|v_0 - \overline{v}\|^2 \beta d}{\mu} \right)$ *and the stochastic approximation scheme is run with step size* $\eta = \frac{1}{c_0 \sigma_L \sqrt{dn}}$, *then the averaged iterates satisfy*

$$\mathbb{E}\|\hat{v}_n - v^*\|_{\mathbb{X}}^2 \leq (1+\omega) \alpha\left(M, 1 - \frac{\mu}{\beta}\right) \inf_{v \in \mathbb{S}} \|v - v^*\|_{\mathbb{X}}^2 + c\left(1 + \frac{1}{\omega}\right) \cdot (\mathcal{E}_n(M, \Sigma_L + \Sigma_b) + \mathcal{H}_n(\sigma_L, \sigma_b, \overline{v}))$$

for any $\omega > 0$.

See Appendix H.3.2 in the online supplementary file for the proof of this corollary.

Note that the approximation factor $\alpha\left(M, 1 - \frac{\mu}{\beta}\right)$ scales as $\mathcal{O}(\beta/\mu)$, which recovers Céa's energy estimates in the symmetric and uniform elliptic case (Céa [16]). On the other hand, for a suitable choice of basis vectors, the bound in Corollary 4 can often be much smaller: the parameter μ corresponding to a global coercive condition can be replaced by the smallest eigenvalue of the *projected* operator M. Furthermore, note that our analysis does not require the symmetry and contraction condition of the operator L, and so it applies also to the case where the operator A is not uniformly elliptic.

It is also worth noting that the bound in Corollary 4 is given in terms of Sobolev norm $\|\cdot\|_{\mathbb{X}} = \|\cdot\|_{\dot{\mathbb{H}}^1}$ as opposed to standard $\mathbb{L}^2$-norm used in the nonparametric estimation literature. By the Poincaré inequality, a Sobolev $\dot{\mathbb{H}}^1$-norm bound implies an $\mathbb{L}^2$-norm bound, and it ensures stronger error guarantees on the gradient of the estimated function.

4.3. Temporal Difference Learning

We now turn to the final example previously introduced in Section 2.2.3—namely, that of the TD algorithm in reinforcement learning. Recall the i.i.d. observation model (21). Also recall the equivalent form of the projected fixed-point equation (23), and note that the population-level operator L satisfies the norm bound

$$|||L|||_{\mathbb{X}} = \gamma \cdot \sup_{\|v\| \leq 1} \|Pv\| \leq \gamma := \gamma_{\max},$$

because ξ is the stationary distribution of the transition kernel P.

4.3.1. Upper Bounds on Stochastic Approximation with Averaging.

As mentioned before, this example is somewhat nonstandard in that the basis functions ψ_i are not necessarily orthonormal; indeed, the classical *temporal difference* learning update in $\mathbb{R}^d$ involves the stochastic approximation algorithm

$$\vartheta_{t+1} = \vartheta_t - \eta(\psi(s_{t+1})\psi(s_{t+1})^\top \vartheta_t - \gamma \psi(s_{t+1})\psi(s_{t+1}^+)^\top \vartheta_t - R_{t+1}(s_{t+1})\psi(s_{t+1})). \tag{41a}$$

The Polyak–Ruppert averaged estimator is then given by the relations

$$\hat{\vartheta}_n = \frac{2}{n} \sum_{t=n/2}^{n-1} \vartheta_t \quad \text{and} \quad \hat{v}_n := \hat{\vartheta}_n^\top \psi. \tag{41b}$$

Note that the updates (22) are, strictly speaking, different from the canonical iterates (25), but this should not be viewed as a fundamental difference because we are ultimately interested in the value function iterates $\hat{v}_n$; these are obtained from the iterates $\hat{\vartheta}_n$ by passing back to the original Hilbert space.

Nevertheless, this cosmetic difference necessitates some natural basis transformations before stating our results. Define the matrix[7] $B \in \mathbb{R}^{d\times d}$ by $B_{ij} := \langle \psi_i, \psi_j \rangle$ for $i, j \in [d]$; this defines an orthonormal basis given by

$$[\,\phi_1 \quad \phi_2 \cdots \phi_d\,] := [\,\psi_1 \quad \psi_2 \cdots \psi_d\,]B^{-1/2}.$$

We define the *min/max eigenvalues* $\beta := \lambda_{\max}(B)$ and $\mu := \lambda_{\min}(B)$ so that β/μ is the condition number of the covariance matrix of the features.

Having set up this transformation, we are now ready to state the implication of our main theorem to the case of LSTD problems. We assume the following fourth-moment condition:

$$\mathbb{E}_\xi[R^4(s)] \le \varsigma^4, \quad \text{and} \quad \mathbb{E}_\xi(u^\top B^{-1/2}\psi(s))^4 \le \varsigma^4 \quad \text{for all } u \in \mathbb{S}^{d-1}. \tag{42}$$

As verified in the proof of Corollary 5 to follow, Equation (42) suffices to guarantee that Assumption 1A is satisfied with parameters $(\sigma_L, \sigma_b) = (2\varsigma^2, \varsigma^2/\sqrt{\beta})$. We also define the matrices

$$M := \gamma B^{-1/2}\mathbb{E}_\xi[\psi(s)\psi(s^+)^\top]B^{-1/2}, \quad \text{and} \quad \Sigma^* := \operatorname{cov}_\xi[B^{-1/2}\psi(s)(\psi(s) - \gamma\psi(s^+) - R(s))^\top].$$

The following corollary then provides a guarantee on the Polyak–Ruppert averaged TD(0) iterates (41).

Corollary 5. *Under the setup above, there are universal positive constants (c, c_0) such that given a sample size $n \ge \frac{c_0\varsigma^4\beta^2 d}{\mu^2(1-\kappa(M))^2}\log^2\left(\frac{\|v_0-\overline{v}\|_2^2\beta d}{\mu(1-\kappa(M))}\right)$, and when the stochastic approximation scheme (41a) is run with step size $\eta = \frac{1}{c_0\varsigma^2\beta\sqrt{dn}}$, the averaged iterates satisfy the bound*

$$\mathbb{E}\|\hat{v}_n - v^*\|^2 \le (1+\omega)\alpha(M,\gamma)\mathcal{A}(\mathbb{S}, v^*) + c\left(1+\frac{1}{\omega}\right)\left[\mathcal{E}_n(M,\Sigma^*) + (1+\|\overline{v}\|^2)\left(\frac{\varsigma^2\beta}{(1-\kappa(M))\mu}\sqrt{\frac{d}{n}}\right)^3\right] \tag{43}$$

for any $\omega > 0$.

See Appendix H.1 in the online supplementary file for the proof of this corollary.

In the worst case, the approximation factor $\alpha(M,\gamma)$ scales as $\frac{1}{1-\gamma^2}$, recovering the classical result (6). More generally, it gives a fine-grained characterization of the approximation factor depending on the one-step autocovariance matrix for the feature vectors. By Lemma 1, we have $\alpha(M,\gamma) \le O\left(\frac{1}{1-\kappa(M)}\right)$, so intuitively, the approximation factor is large when there are feature space directions in which the Markov chain transitions slowly. On the other hand, if the one-step transitions move rapidly in all directions of feature space, then the approximation factor is much smaller.

The statistical error term $\mathcal{E}_n(M,\Sigma^*)$ matches the Cramér–Rao lower bound and gives a finer characterization than both worst-case upper bounds (Bhandari et al. [8]) and existing instance-dependent upper bounds (Lakshminarayanan and Szepesvári [33]). Note that the final higher-order term depends on the condition number $\frac{\beta}{\mu}$ of the covariance matrix B. This ratio is 1 when the basis vectors are orthonormal, but in general, the speed of algorithmic convergence depends on this parameter.

4.3.2. Approximation Factor Lower Bounds for MRPs. We conclude our discussion of discounted MRPs with an information-theoretic lower bound for policy evaluation. This bound involves technical effort beyond that in the proof of Theorem 2, because any valid construction for MRPs must make use only of operators L that are constructed using a valid transition kernel. To set the stage, we say that a Markov reward process (P,γ,r) and associated basis functions $\{\psi_j\}_{j=1}^d$ are in the *canonical setup* if the following conditions hold:

- The stationary distribution ξ of P exists and is unique.
- The reward function and its observations are uniformly bounded. In particular, we have $\|r\|_\infty \le 1$, and $\|R\|_\infty \le 1$ almost surely.
- The basis functions are orthonormal; that is, $\mathbb{E}_\xi[\psi(s)\psi(s)^\top] = I_d$.

The three conditions are standard assumptions in Markov reward processes.

Now given scalars $\nu \in (0,1]$ and $\gamma \in (0,1)$, integer $D > 0$, and scalar $\delta \in (0,1/2)$, we consider the following class of MRPs and associated feature vectors:

$$\mathbb{C}_{\mathsf{MRP}}(\nu,\gamma,D,\delta) := \left\{(P,\gamma,r,\psi) \;\middle|\; \begin{array}{c}(P,\gamma,r,\psi) \text{ is in the canonical setup}, \quad |\mathcal{S}| = D, \\ \mathcal{A}(\mathbb{S},v^*) \le \delta^2, \quad \kappa(\mathbb{E}_\xi[\psi(s)\psi(s^+)^\top]) \le \nu\end{array}\right\}.$$

Note that under the canonical setup, we have $M = \gamma \mathbb{E}_\xi[\psi(s)\psi(s^+)^\top]$, and consequently, a problem instance in the class $\mathbb{C}_{\mathsf{MRP}}(\nu, \gamma, D, \delta)$ satisfies $\kappa(M) \le \nu\gamma$ in the setup of Corollary 5. The condition $\kappa(\mathbb{E}_\xi[\psi(s)\psi(s^+)^\top]) \le \nu$ can be seen as a "mixing" condition in the projected space: when ν is bounded away from 1, the feature vector cannot have too large a correlation with its next-step transition in any direction.

We have the following minimax lower bound for this class, where we use the shorthand $\mathbb{C}_{\mathsf{MRP}} \equiv \mathbb{C}_{\mathsf{MRP}}(\nu, \gamma, D, \delta)$ for convenience.

Proposition 1. *There are universal positive constants (c, c_1) such that if $D \ge c_1(n^2 + d)$, then for all scalars $\nu \in (0,1]$ and $\gamma \in (0,1)$, we have*

$$\inf_{\hat{v}_n \in \hat{\mathbb{V}}_{\mathbb{X}}} \sup_{(P,\gamma,r,\psi) \in \mathbb{C}_{\mathsf{MRP}}} \|\hat{v}_n - v^*\|^2 \ge \frac{c}{1-\nu\gamma}\delta^2 \wedge 1. \tag{44}$$

See Appendix H.2 in the online supplementary file for the proof of this proposition.

A few remarks are in order. First, in conjunction with Corollary 5 and the second upper bound in Lemma 1, we can conclude that the TD algorithm for policy evaluation with linear function approximation attains the minimax-optimal approximation factor over the class $\mathbb{C}_{\mathsf{MRP}}$ up to universal constants, in the regime where the optimal error is bounded by $O(1)$. It is also worth noting that Proposition 1 also shows that the worst-case upper bound (6) attributable to Tsitsiklis and Van Roy [60] is indeed sharp up to a universal constant; indeed, note that for all $\gamma \in (0,1)$, we have $\frac{1}{1-\gamma^2} \asymp \frac{1}{1-\gamma}$ and that the latter factor can be obtained from the lower bound (44) by taking $\nu = 1$.

Second, note that the class $\mathbb{C}_{\mathsf{MRP}}$ is defined in a more "global" sense as opposed to the "local" class $\mathbb{C}_{\mathsf{approx}}$ used in Theorem 2. This class contains all the MRP instances satisfying the approximation error bound and the constraint on $\kappa(M)$, and a minimax lower bound over this larger class is weaker than the lower bound over the local class that imposes restrictions on the projected matrix. That being said, Proposition 1 still captures more structure in the Markov transition kernel than the fact that it is contractive in the ξ-norm. For example, when the Markov chain makes "local moves" in the feature space, the correlation between feature vectors can be large, leading to a large value of ν and larger values of the optimal approximation factor. On the other hand, if the one-step transition of the feature vector jumps a large distance in all directions, the optimal approximation factor will be small.

Finally, it is worth noticing that Proposition 1 holds only for the i.i.d. observation models. If we are given the entire trajectory of the Markov reward process, the approximation factor can be made arbitrarily close to 1, using TD(λ) methods (Tsitsiklis and Van Roy [60]). The trade-off inherent to the Markov observation model is an important direction for future work.

5. Proofs

We now turn to the proofs of our main results. The main proofs of Theorems 1 and 2 are given in this section, with some technical lemmas deferred to the online supplementary file. The proofs of Theorems 3 and 4 and Corollaries 1 and 2, as well as associated lemmas, are presented in the online supplementary file.

5.1. Proof of Theorem 1

We divide the proof into two parts, corresponding to the two components in the mean-squared error of the estimator $\hat{v}_n$. The first term is the *approximation error* $\|\overline{v} - v^*\|^2$ that arises from the difference between the exact solution v^* to the original fixed-point equation and the exact solution $\overline{v}$ to the projected set of equations. The second term is the *estimation error* $\mathbb{E}\|\hat{v}_n - \overline{v}\|^2$, measuring the difficulty of estimating $\overline{v}$ on the basis of n noisy samples.

In particular, under the conditions of the theorem, we prove that the approximation error is upper bounded as

$$\|\overline{v} - v^*\|^2 \le \alpha(M, |||L|||_{\mathbb{X}}) \inf_{v \in \mathbb{S}} \|v - v^*\|^2, \tag{45a}$$

whereas the estimation error is bounded as

$$\mathbb{E}\|\hat{v}_n - \overline{v}\|^2 \le c\frac{\mathrm{trace}((I-M)^{-1}\Sigma^*(I-M)^{-\top})}{n} + c\frac{\sigma_L}{(1-\kappa)^3}\left(\frac{d}{n}\right)^{3/2}(\|\overline{v}\|^2\sigma_L^2 + \sigma_b^2). \tag{45b}$$

Given these two inequalities, it is straightforward to prove the bound (28) stated in the theorem. By expanding the square, we have

$$\begin{aligned}\mathbb{E}\|\hat{v}_n - v^*\|^2 &= \mathbb{E}\|\hat{v}_n - \overline{v}\|^2 + \|\overline{v} - v^*\|^2 + 2\mathbb{E}\langle \hat{v}_n - \overline{v}, \overline{v} - v^*\rangle \\ &\overset{\text{(i)}}{\leq} \mathbb{E}\|\hat{v}_n - \overline{v}\|^2 + \|\overline{v} - v^*\|^2 + 2\sqrt{\mathbb{E}\|\hat{v}_n - \overline{v}\|^2 \cdot \|\overline{v} - v^*\|^2} \\ &\overset{\text{(ii)}}{\leq} \mathbb{E}\|\hat{v}_n - \overline{v}\|^2 + \|\overline{v} - v^*\|^2 + \frac{1}{\omega}\mathbb{E}\|\hat{v}_n - \overline{v}\|^2 + \omega\|\overline{v} - v^*\|^2 \\ &= (1+\omega)\|\overline{v} - v^*\|^2 + \left(1 + \frac{1}{\omega}\right)\mathbb{E}\|\hat{v}_n - \overline{v}\|^2,\end{aligned}$$

where step (i) follows from the Cauchy–Schwarz inequality, and step (ii) follows from the arithmetic-geometric mean inequality and is valid for any $\omega > 0$. Substituting the bounds from Equations (45a) and (45b) yields the claim of the theorem.

The remainder of our argument is devoted to the proofs of the bounds (45a) and (45b).

5.1.1. Proof of Approximation Error Bound (45a). We begin with some decomposition relations for vectors and operators. Note that $\mathbb{S}$ is a finite-dimensional subspace and therefore is closed. We use

$$\mathbb{S}^\perp := \{u \in \mathbb{X} \mid \langle u, v\rangle = 0 \mid \text{for all } v \in \mathbb{S}\}$$

to denote its orthogonal complement. The pair $(\mathbb{S}, \mathbb{S}^\perp)$ forms a direct product decomposition of $\mathbb{X}$, and the projection operators satisfy $\Pi_{\mathbb{S}} + \Pi_{\mathbb{S}^\perp} = I$. Also define the operators $L_{\mathbb{S},\mathbb{S}} = \Pi_{\mathbb{S}} L \Pi_{\mathbb{S}}$ and $L_{\mathbb{S},\perp} = \Pi_{\mathbb{S}} L \Pi_{\mathbb{S}^\perp}$. With this notation, our proof can be broken down into two auxiliary lemmas, which we state here.

Lemma 3. *The error $\|\overline{v} - v^*\|$ between the projected fixed point $\overline{v}$ and the original fixed point v^* is bounded as*

$$\|\overline{v} - v^*\|^2 \leq (1 + |||(I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp}|||_{\mathbb{X}}^2) \inf_{v \in \mathbb{S}} \|v - v^*\|^2. \tag{46}$$

Lemma 4. *Under the above-mentioned setup, we have*

$$|||(I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp}|||_{\mathbb{X}}^2 \leq \lambda_{\max}((I_d - M)^{-1}(|||L|||_{\mathbb{X}}^2 I_d - MM^\top)(I_d - M)^{-\top}).$$

The claimed bound (45a) on the approximation error follows by combining these two lemmas and recalling our definition of $\alpha(M, L)$. We now prove these two lemmas in turn.

5.1.2. Proof of Lemma 3. For any vector $v \in \mathbb{X}$, we perform the orthogonal decomposition $v = v_{\mathbb{S}} + v_\perp$, where $v_{\mathbb{S}} := \Pi_{\mathbb{S}}(v)$ is a member of the set $\mathbb{S}$, and $v_\perp := \Pi_{\mathbb{S}^\perp, \xi}$ is a member of the set $\mathbb{S}^\perp$. With this notation, the operator L can be decomposed as

$$L = (\Pi_{\mathbb{S}} + \Pi_{\mathbb{S}^\perp}) L (\Pi_{\mathbb{S}} + \Pi_{\mathbb{S}^\perp}) = \underbrace{\Pi_{\mathbb{S}} L \Pi_{\mathbb{S}}}_{=:L_{\mathbb{S},\mathbb{S}}} + \underbrace{\Pi_{\mathbb{S}} L \Pi_{\mathbb{S}^\perp}}_{=:L_{\mathbb{S},\perp}} + \underbrace{\Pi_{\mathbb{S}^\perp} L \Pi_{\mathbb{S}}}_{=:L_{\perp,\mathbb{S}}} + \underbrace{\Pi_{\mathbb{S}^\perp} L \Pi_{\mathbb{S}^\perp}}_{=:L_{\perp,\perp}}.$$

The four operators $L_{\mathbb{S},\mathbb{S}}, L_{\mathbb{S},\perp}, L_{\perp,\mathbb{S}}$, and $L_{\perp,\perp}$ defined in the preceding equation are also bounded linear operators. By the properties of projection operators, we note that $L_{\mathbb{S},\mathbb{S}}$ and $L_{\perp,\mathbb{S}}$ both map each element of $\mathbb{S}^\perp$ to 0, and $L_{\mathbb{S},\perp}$ and $L_{\perp,\perp}$ both map each element of $\mathbb{S}$ to 0.

Decomposing the target vector v^* in an analogous manner yields the two components

$$\tilde{v} := \Pi_{\mathbb{S}}(v^*) \quad \text{and} \quad v^\perp := v^* - \tilde{v}.$$

The fixed-point equation $v^* = Lv^* + b$ can then be written using $\mathbb{S}$ and its orthogonal complement as

$$\tilde{v} \overset{(a)}{=} L_{\mathbb{S},\mathbb{S}}\tilde{v} + L_{\mathbb{S},\perp}v^\perp + b_{\mathbb{S}}, \quad \text{and} \quad v^\perp \overset{(b)}{=} L_{\perp,\mathbb{S}}\tilde{v} + L_{\perp,\perp}v^\perp + b_\perp. \tag{47}$$

For the projected solution $\overline{v}$, we have the defining equation

$$\overline{v} = L_{\mathbb{S},\mathbb{S}}\overline{v} + b_{\mathbb{S}}. \tag{48}$$

Subtracting (a) in Equation (47) from Equation (48) yields

$$(I - L_{\mathbb{S},\mathbb{S}})(\tilde{v} - \overline{v}) = L_{\mathbb{S},\perp}v^\perp.$$

Recall the quantity $M = \Phi_d L \Phi_d^*$ and our assumption that $\kappa(M) = \frac{1}{2}\lambda_{\max}(M + M^T) < 1$. This condition implies that $I - L_{\mathbb{S},\mathbb{S}}$ is invertible on the subspace $\mathbb{S}$. Because this operator also maps each element of $\mathbb{S}^\perp$ to itself, it is invertible on all of $\mathbb{X}$, and we have $\tilde{v} - \overline{v} = (I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp} v^\perp$.

Applying the Pythagorean theorem then yields

$$\begin{aligned}\|\overline{v} - v^*\|^2 &= \|\overline{v} - \tilde{v}\|^2 + \|\tilde{v} - v^*\|^2 = \|(I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp} v^\perp\|^2 + \|v^\perp\|^2 \\ &\leq (1 + |||(I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp}|||_{\mathbb{X}}^2) \cdot \|v^\perp\|^2, \end{aligned} \tag{49}$$

as claimed.

5.1.3. Proof of Lemma 4. By the definition of operator norm for any vector $v \in \mathbb{X}$ such that $\|v\| = 1$, we have

$$|||L|||_{\mathbb{X}}^2 \geq \|Lv\|^2 = \|L_{\mathbb{S},\mathbb{S}} v_{\mathbb{S}} + L_{\mathbb{S},\perp} v_\perp\|^2 + \|L_{\perp,\mathbb{S}} v_{\mathbb{S}} + L_{\perp,\perp} v_\perp\|^2 \geq \|L_{\mathbb{S},\mathbb{S}} v_{\mathbb{S}} + L_{\mathbb{S},\perp} v_\perp\|^2.$$

Noting the fact that $L_{\mathbb{S},\mathbb{S}} v_\perp = 0 = L_{\mathbb{S},\perp} v_{\mathbb{S}}$, we have the following norm bound on the linear operator $L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp}$:

$$\begin{aligned}|||L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp}|||_{\mathbb{X}} &= \sup_{\|v\|=1} \|(L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp})v\| \\ &= \sup_{\|v\|=1} \|L_{\mathbb{S},\mathbb{S}} v_{\mathbb{S}} + L_{\mathbb{S},\perp} v_\perp\| \leq |||L|||_{\mathbb{X}}. \end{aligned}$$

By definition, the operator $L^*_{\mathbb{S},\perp} = \Pi_{\mathbb{S}^\perp} L^* \Pi_{\mathbb{S}}$ maps any vector to $\mathbb{S}^\perp$, and the operator $L_{\mathbb{S},\mathbb{S}}$ maps any element of $\mathbb{S}^\perp$ to 0. Therefore, we have the identity $L_{\mathbb{S},\mathbb{S}} L^*_{\mathbb{S},\perp} = 0$. A similar argument yields that $L_{\mathbb{S},\perp} L^*_{\mathbb{S},\mathbb{S}} = 0$. Consequently, we have

$$\begin{aligned}|||L|||_{\mathbb{X}}^2 \geq |||L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp}|||_{\mathbb{X}}^2 &= |||(L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp})(L_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp})^*|||_{\mathbb{X}} \\ &= |||\underbrace{L_{\mathbb{S},\mathbb{S}} L^*_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp} L^*_{\mathbb{S},\perp}}_{=:G}|||_{\mathbb{X}}. \end{aligned} \tag{50}$$

Note that the operator G can be expressed as $G = \Pi_{\mathbb{S}}(L\Pi_{\mathbb{S}} L^* + L\Pi_{\mathbb{S}^\perp} L^*)\Pi_{\mathbb{S}}$. From this representation, we see that

- for any vector $x \in \mathbb{X}$, we have $Gx \in \mathbb{S}$, and
- for any vector $y \in \mathbb{S}^\perp$, we have $Gy = 0$.

Consequently, there exists a matrix $\tilde{G} \in \mathbb{R}^{d \times d}$ such that $G = \Phi_d^* \tilde{G} \Phi_d$. Because G is a positive semidefinite operator, the matrix $\tilde{G}$ is positive semidefinite. Equation (50) implies that

$$\lambda_{\max}(\tilde{G}) = |||\tilde{G}|||_{\text{op}} = |||G|||_{\mathbb{X}} \leq |||L|||_{\mathbb{X}}^2. \tag{51a}$$

Now defining $\tau := |||(I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp}|||_{\mathbb{X}}$, note that

$$\tau^2 = |||\underbrace{(I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp} L^*_{\mathbb{S},\perp} (I - L^*_{\mathbb{S},\mathbb{S}})^{-1}}_{=:H}|||_{\mathbb{X}}. \tag{51b}$$

Moreover, the operator H is self-adjoint, and we have the following properties:

- The operator $L_{\mathbb{S},\perp}$ maps any vector to $\mathbb{S}$, and $(I - L_{\mathbb{S},\mathbb{S}})^{-1}$ maps $\mathbb{S}$ to itself. Consequently, for any $x \in \mathbb{X}$, the vector $Hx = (I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp} (L^*_{\mathbb{S},\perp} (I - L^*_{\mathbb{S},\mathbb{S}})^{-1}) x$ is a member of the set $\mathbb{S}$.
- The operator $L^*_{\mathbb{S},\perp} = \Pi_{\mathbb{S}^\perp} L^* \Pi_{\mathbb{S}}$ maps any vector from $\mathbb{S}^\perp$ to 0. Consequently, for any $y \in \mathbb{S}^\perp$, we have $Hy = (I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp} (L^*_{\mathbb{S},\perp} (I - L^*_{\mathbb{S},\mathbb{S}})^{-1}) y = 0$.

Given these facts, there exists a matrix $\tilde{H} \in \mathbb{R}^{d \times d}$ such that $H = \Phi_d^* \tilde{H} \Phi_d$. Because the operator H is positive semidefinite, so too is the matrix $\tilde{H}$. Consequently, by Equation (51b), we obtain the identity $\tau^2 = |||H|||_{\mathbb{X}} = |||\tilde{H}|||_{\text{op}} = \lambda_{\max}(H)$. In particular, letting $u \in \mathbb{S}^{d-1}$ be a maximal eigenvector of $\tilde{H}$, we have

$$\tilde{H} \geq \tau^2 u u^\top. \tag{52}$$

Because $M = \Phi_d L_{\mathbb{S},\mathbb{S}} \Phi_d^*$ by definition, combining the aforementioned matrix inequalities (51a) and (52), we arrive at the bound

$$\begin{aligned}|||L|||_{\mathbb{X}}^2 I_d &\geq \tilde{G} \\ &= \Phi_d (L_{\mathbb{S},\mathbb{S}} L^*_{\mathbb{S},\mathbb{S}} + L_{\mathbb{S},\perp} L^*_{\mathbb{S},\perp}) \Phi_d^* \\ &= \Phi_d L_{\mathbb{S},\mathbb{S}} L^*_{\mathbb{S},\mathbb{S}} \Phi_d^* + (\Phi_d (I - L_{\mathbb{S},\mathbb{S}}) \Phi_d^*) \cdot (\Phi_d (I - L_{\mathbb{S},\mathbb{S}})^{-1} L_{\mathbb{S},\perp} L^*_{\mathbb{S},\perp} (I - L^*_{\mathbb{S},\mathbb{S}})^{-1} \Phi_d^*) \cdot (\Phi_d (I - L^*_{\mathbb{S},\mathbb{S}}) \Phi_d^*) \\ &= MM^\top + (I - M) \tilde{H} (I - M^\top) \\ &\geq MM^\top + \tau^2 (I - M) u u^\top (I - M^\top). \end{aligned}$$

Rearranging and noting that $u \in \mathbb{S}^{d-1}$, we arrive at the inequality

$$\tau^2 \leq u^\top [(I - M)^{-1} (|||L|||_{\mathbb{X}}^2 I_d - MM^\top)(I - M)^{-\top}] u \leq \lambda_{\max}((I - M)^{-1} (|||L|||_{\mathbb{X}}^2 I_d - MM^\top)(I - M)^{-\top}),$$

which completes the proof of Lemma 4.

5.1.4. Proof of Estimation Error Bound (45b). We now turn to the proof of our claimed bound on the estimation error. Our analysis relies on two auxiliary lemmas. The first lemma provides bounds on the mean-squared error of the standard iterates $\{v_t\}_{t\geq 0}$—that is, without the averaging step.

Lemma 5. *Suppose that the noise conditions in Assumption 1A hold. Then for any step size* $\eta \in \big(0, (1-\kappa)/(4\sigma_L^2 d + 1 + |||L|||_{\mathbb{X}}^2)\big)$, *we have the bound*

$$\mathbb{E}\|v_t - \overline{v}\|^2 \leq e^{-(1-\kappa)\eta t/2}\mathbb{E}\|v_0 - \overline{v}\|^2 + \frac{8\eta}{1-\kappa}(\|\overline{v}\|^2\sigma_L^2 d + \sigma_b^2 d) \qquad \text{valid for } t = 1, 2, \ldots. \tag{53}$$

See Appendix A.1 in the online supplementary file for the proof of this claim.

Our second lemma provides a bound on the Polyak–Ruppret-averaged estimate $\hat{v}_n$ based on n observations in terms of a covariance term, along with the error of the nonaveraged sequences $\{v_t\}_{t\geq 1}$.

Lemma 6. *Under the above-mentioned setup, we have the bound*

$$\begin{aligned}\mathbb{E}\|\hat{v}_n - \overline{v}\|^2 \leq{}& \frac{3}{n-n_0}\operatorname{trace}((I-M)^{-1}\Sigma^*(I-M)^{-\top}) \\ &+ \frac{3}{(n-n_0)^2}\sum_{t=n_0}^{n}\mathbb{E}\|(I-M)^{-1}\Phi_d(L_{t+1}-L)(v_t-\overline{v})\|_2^2 + \frac{3\mathbb{E}\|v_n - v_{n_0}\|^2}{\eta^2(n-n_0)^2(1-\kappa)^2}.\end{aligned} \tag{54}$$

See Appendix A.2 in the online supplementary file for the proof of this claim.

Equipped with these two lemmas, we can now complete the proof of the claimed bound (45b) on the estimation error. Recalling that $n_0 = n/2$, we see that the first term in the bound (54) matches a term in the bound (45b). As for the remaining two terms in Equation (54), the second-moment bounds from Assumption 1A combined with the assumption that $\kappa(M) < 1$ imply that

$$\begin{aligned}\mathbb{E}\|(I-M)^{-1}\Phi_d(L_{t+1}-L)(v_t-\overline{v})\|_2^2 &\leq \frac{1}{(1-\kappa)^2}\mathbb{E}\|\Phi_d(L_{t+1}-L)(v_t-\overline{v})\|_2^2 \\ &\leq \frac{1}{(1-\kappa)^2}\sum_{j=1}^{d}\mathbb{E}\langle\phi_j, (L_{t+1}-L)(v_t-\overline{v})\rangle^2 \leq \frac{\sigma_L^2 d\|v_t-\overline{v}\|^2}{(1-\kappa)^2}.\end{aligned}$$

On the other hand, we can use Lemma 5 to control the third term in the bound (54). We begin by observing that

$$\|v_n - v_{n_0}\|^2 \leq 2\|v_n - \overline{v}\|^2 + 2\|v_{n_0} - \overline{v}\|^2 \leq 4\sup_{n_0\leq t\leq n}\mathbb{E}\|v_t - \overline{v}\|^2.$$

If we choose a burn-in time $n_0 > \frac{c_0}{(1-\kappa)\eta}\log\left(\frac{\|v_0-\overline{v}\|^2 d}{1-\kappa}\right)$, then Lemma 5 ensures that

$$\sup_{n_0\leq t\leq n}\mathbb{E}\|v_t - \overline{v}\|^2 \leq \frac{16\eta}{1-\kappa}(\|\overline{v}\|^2\sigma_L^2 d + \sigma_b^2 d).$$

Finally, taking the step size $\eta = (24\sigma_L\sqrt{dn})^{-1}$, recalling that $n_0 = n/2$, and putting together the pieces yields

$$\begin{aligned}\mathbb{E}\|\hat{v}_n - \overline{v}\|^2 &\leq \frac{12}{n}\operatorname{trace}((I-M)^{-1}\Sigma^*(I-M)^{-\top}) + \frac{1}{(1-\kappa)^2}\left(\frac{12\sigma_L^2 d}{n} + \frac{48}{\eta^2 n^2}\right)\sup_{n_0\leq t\leq n}\mathbb{E}\|v_t - \overline{v}\|^2 \\ &\leq \frac{12}{n}\operatorname{trace}((I-M)^{-1}\Sigma^*(I-M)^{-\top}) + \frac{48\sigma_L}{(1-\kappa)^3}\left(\frac{d}{n}\right)^{3/2}(\|\overline{v}\|^2\sigma_L^2 + \sigma_b^2),\end{aligned}$$

as claimed.

5.2. Proof of Theorem 2

At a high level, our proof of the lower bound proceeds by constructing two ensembles of problem instances that are hard to distinguish from each other and such that the approximation error on at least one of them is large. The two instances are indexed by values of a bit $z \in \{-1, 1\}$, and each instance is, in turn, obtained as a mixture over 2^{D-d} centers; each center is indexed by a binary string $\varepsilon \in \{-1, 1\}^{D-d}$. The problem is then phrased as one of estimating the value of z from the observations; this is effectively a reduction to testing and the use of Le Cam's mixture-versus-mixture method.

Specifically, let $u \in \mathbb{S}^{d-1}$ be an eigenvector associated with the largest eigenvalue of the matrix $(I-M_0)^{-1}(\gamma_{\max}^2 I - M_0 M_0^{\top})(I-M_0)^{-\top}$. By the definition of the approximation factor $\alpha(M_0, \gamma_{\max})$, we have

$$(\alpha(M_0, \gamma_{\max}) - 1) \cdot (I - M_0) u u^\top (I - M_0)^\top \preceq \gamma_{\max}^2 I - M_0 M_0^\top.$$

Based on the eigenvector u, we further define the d-dimensional vectors:

$$w := \sqrt{\alpha(M_0, \gamma_{\max}) - 1} \cdot (I - M_0) u \quad \text{and} \quad y := \sqrt{\alpha(M_0, \gamma_{\max}) - 1} \cdot \delta u. \tag{55}$$

Substituting into the aforementioned PSD domination relation yields that

$$w w^\top + M_0 M_0^\top \preceq \gamma_{\max}^2 I. \tag{56}$$

Now consider the following class of (population-level) problem instances $(L^{(\varepsilon,z)}, b^{(\varepsilon,z)}, v^*_{\varepsilon,z})$ indexed by a binary string $\varepsilon \in \{-1,1\}^{D-d}$ and a bit $z \in \{-1,1\}$:

$$L^{(\varepsilon,z)} := \begin{bmatrix} M_0 & \frac{\sqrt{d}}{D-d}\varepsilon_{d+1} w & \cdots & \frac{\sqrt{d}}{D-d}\varepsilon_D w \\ 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \; v^*_{\varepsilon,z} := \begin{bmatrix} \sqrt{2d}(zy + (I - M_0)^{-1} h_0) \\ \sqrt{2} z \delta \varepsilon_{d+1} \\ \vdots \\ \sqrt{2} z \delta \varepsilon_D \end{bmatrix},$$

$$b^{(\varepsilon,z)} := (I - L^{(\varepsilon,z)}) v^*_{\varepsilon,z} = \begin{bmatrix} \sqrt{2d} h_0 \\ \sqrt{2} z \delta \varepsilon_{d+1} \\ \vdots \\ \sqrt{2} z \delta \varepsilon_D \end{bmatrix}. \tag{57}$$

We take the weight vector ξ to be

$$\xi = \left[\underbrace{\frac{1}{2d} \cdots \frac{1}{2d}}_{d} \; \underbrace{\frac{1}{2(D-d)} \cdots \frac{1}{2(D-d)}}_{(D-d)} \right],$$

and the weighted inner product $\langle \cdot, \cdot \rangle$ on the space $\mathbb{X} = \mathbb{R}^D$ is defined via

$$\langle p, q \rangle := \sum_{j=1}^D p_j \xi_j q_j \quad \text{for each pair } p, q \in \mathbb{R}^D.$$

This choice of inner product then induces the vector norm $\|\cdot\|$ and operator norm $|||\cdot|||_{\mathbb{X}}$.

Next, we define the basis vectors via

$$\phi_i = \begin{cases} \sqrt{2d} e_i & \text{for } i = 1, 2, \ldots, d, \text{ and} \\ \sqrt{2(D-d)} e_i & \text{for } i = d+1, \ldots, D. \end{cases}$$

By construction, we have ensured that $\|\phi_i\| = 1$ for each $i \in [D]$. We let the subspace $\mathbb{S}$ be the span of the first d standard basis vectors (i.e., $\mathbb{S} := \text{span}(e_1, e_2, \ldots, e_d)$).

For each binary string $\varepsilon \in \{-1,1\}^{D-d}$ and signed bit $z \in \{-1,1\}$, a straightforward calculation reveals that the projected problem instance satisfies the identities

$$\Phi_d L^{(\varepsilon,z)} \Phi_d^* = M_0 \quad \text{and} \quad \Phi_d b^{(\varepsilon,z)} = h_0. \tag{58a}$$

Also note that for any pair (ε, z), we have by construction that

$$\inf_{v \in \mathbb{S}} \|v^*_{\varepsilon,z} - v\|^2 = \frac{1}{2(D-d)} \sum_{j=d+1}^D (\sqrt{2} z \delta \varepsilon_j)^2 = \delta^2. \tag{58b}$$

In other words, this shows that the $\|\cdot\|$-error of approximating $v^*_{\varepsilon,z}$ with the linear subspace $\mathbb{S}$ is always δ, irrespective of which $\varepsilon \in \{-1,1\}^{D-d}$ and $z \in \{-1,1\}$ are chosen.

Next, we construct the random observation models for the i.i.d. observations, which are also indexed by the pair (ϵ, z). In particular, we construct the random matrix $L_i^{(\varepsilon,z)}$ and random vector $b_i^{(\varepsilon,z)}$ via

$$L_i^{(\varepsilon,z)} := \begin{bmatrix} M_0 & 0 & \cdots & 0 & \sqrt{d}\varepsilon_{\tau_L^{(i)}} w & 0 & \cdots & 0 \\ 0 & 0 & & & \cdots & & & 0 \\ & & & \vdots & & & \vdots & \\ 0 & 0 & & & \cdots & & & 0 \end{bmatrix}, \quad b_i^{(\varepsilon,z)} := \begin{bmatrix} \sqrt{2d}h_0 \\ 0 \\ \vdots \\ 0 \\ \sqrt{2}(D-d)z\delta\varepsilon_{\tau_b^{(i)}} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{59}$$

where the random indices $\tau_L^{(i)}$ and $\tau_b^{(i)}$ are chosen independently and uniformly at random from the set $\{d+1, d+2, \ldots, D\}$. By construction, we have ensured that for each $\varepsilon \in \{-1,1\}^{D-d}$ and $z \in \{-1,1\}$, the observations have means

$$\mathbb{E}[L_i^{(\varepsilon,z)}] = L^{(\varepsilon,z)} \quad \text{and} \quad \mathbb{E}[b_i^{(\varepsilon,z)}] = b^{(\varepsilon,z)}.$$

This concludes our description of the problem instances themselves. Because our proof proceeds via Le Cam's lemma, we require some more notation for product distributions and mixtures under this observation model. Let $\mathbb{P}_{\varepsilon,z}^{(n)}$ denote the n-fold product of the probability laws of the pair $(L_i^{(\varepsilon,z)}, b_i^{(\varepsilon,z)})$. We also define the following mixture of product measures for each $z \in \{-1,1\}$:

$$\mathbb{P}_z^{(n)} := \frac{1}{2^{D-d}} \sum_{\varepsilon \in \{\pm 1\}^{D-d}} \mathbb{P}_{\varepsilon,z}^{(n)}.$$

We seek bounds on the total variation distance $d_{\mathrm{TV}}(\mathbb{P}_1^{(n)}, \mathbb{P}_{-1}^{(n)})$.

With this setup, the following lemmas assert that (a) our construction satisfies the conditions in Assumption 1B, and (b) the total variation distance is small provided $n \lesssim \sqrt{D-d}$.

Lemma 7. *For each binary string $\varepsilon \in \{-1,1\}^{D-d}$ and bit $z \in \{-1,1\}$, we have the following.*

(a) *The population-level matrix $L^{(\varepsilon,z)}$ defined in* Equation (57) *satisfies $|||L^{(\varepsilon,z)}|||_{\mathbb{X}} \le \gamma_{\max}$.*

(b) *The random observations $(L_i^{(\varepsilon,z)}, b_i^{(\varepsilon,z)})$ defined in* Equation (59) *satisfy Assumption 1B for any scalar pair (σ_L, σ_b) such that $\sigma_L \ge \gamma_{\max}$ and $\sigma_b \ge \delta$.*

Lemma 8. *Under the above-mentioned setup, we have $d_{\mathrm{TV}}(\mathbb{P}_1^{(n)}, \mathbb{P}_{-1}^{(n)}) \le \frac{12n^2}{D-d}$.*

See Appendices B.1 and B.2 in the online supplementary file for the proofs of Lemmas 7 and 8, respectively.

Part (a) of Lemma 7 and Equations (58a) and (58b) together ensure that population-level problem instance (L, b) we constructed belongs to the class $\mathbb{C}_{\mathsf{approx}}(M_0, h_0, D, \delta, \gamma_{\max})$. Part (b) of Lemma 7 further ensures the probability distribution $\mathbb{P}_{L,b}$ belongs to the class $\mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)$. Lemma 8 ensures that the two mixture distributions corresponding to different choices of the bit z are close, provided n is not too large. The final step in applying Le Cam's mixture-versus-mixture result is to show that the approximation error is large for at least one of the choices of the bit z. We carry out this step by splitting the rest of the proof into two cases, depending on whether we enforce that our estimator $\hat{v}$ is constrained to lie in the subspace $\mathbb{S}$. Throughout, we use the decomposition $\hat{v} = \begin{bmatrix} \hat{v}_1 \\ \hat{v}_2 \end{bmatrix}$, where $\hat{v}_1 \in \mathbb{R}^d$ and $\hat{v}_2 \in \mathbb{R}^{D-d}$. Also recall the definition of the vector y from Equation (55).

Case 1: $\hat{v} \in \mathbb{S}$.This corresponds to the "proper learning" case where the estimator is restricted to take values in the subspace $\mathbb{S}$ and $\hat{v}_2 = 0$. Note that for any $\varepsilon \in \{-1,1\}^{D-d}$, we have

$$\|v_{\varepsilon,z}^* - \hat{v}\|^2 = \|v_{\varepsilon,z}^* - \Pi_{\mathbb{S}}(v_{\varepsilon,z}^*)\|^2 + \|v_{\varepsilon,z}^* - \hat{v}\|^2 = \delta^2 + \frac{1}{2d}\|\hat{v}_1 - \sqrt{2d}zy\|_2^2.$$

Therefore, for any $\varepsilon, \varepsilon' \in \{-1,1\}^{D-d}$, the following chain of inequalities holds:

$$\begin{aligned} \frac{1}{2}(\|v_{\varepsilon,1}^* - \hat{v}\|^2 + \|v_{\varepsilon',-1}^* - \hat{v}\|^2) &= \delta^2 + \frac{1}{4d}(\|\hat{v}_1 - \sqrt{2d}y\|_2^2 + \|\hat{v}_1 + \sqrt{2d}y\|_2^2) \\ &= \delta^2 + \frac{1}{2d}(\|\hat{v}_1\|_2^2 + 2d\|y\|_2^2) \\ &\ge \delta^2 + \|y\|_2^2 = \alpha(M_0, \gamma_{\max}) \cdot \delta^2. \end{aligned}$$

By Le Cam's lemma, we thus have

$$\inf_{\hat{v}_n \in \hat{\mathcal{V}}_{\mathbb{S}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\text{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\text{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\hat{v}_n - v^*\|^2 \geq \alpha(M_0, \gamma_{\max})\delta^2 \cdot (1 - d_{\text{TV}}(\mathbb{P}_{-1}^{(n)}, \mathbb{P}_1^{(n)})) \overset{\text{(i)}}{\geq} (1-\omega) \cdot \alpha(M_0, \gamma_{\max}) \cdot \delta^2,$$

where in step (i), we have applied Lemma 8 in conjunction with the inequality $D \geq d + \frac{12n^2}{\omega}$.

Case 2: $\hat{v} \notin \mathbb{S}$.This corresponds to the case of "improper learning" where the estimator can take values in the entire space $\mathbb{X}$. In this case, for any pair $\varepsilon, \varepsilon' \in \{-1,1\}^{D-d}$, we obtain

$$\|v^*_{\varepsilon,1} - v^*_{\varepsilon',-1}\| \geq \|[\, 2\sqrt{2d}y^\top \quad 0 \cdots 0\,]^\top\| = 2\|y\|_2 = 2\delta\sqrt{\alpha(M_0, \gamma_{\max}) - 1}.$$

Applying triangle inequality and Young's inequality yields the bound

$$\frac{1}{2}(\|\hat{v} - v^*_{\varepsilon,1}\|^2 + \|\hat{v} - v^*_{\varepsilon',1}\|^2) \geq \frac{1}{4}(\|\hat{v} - v^*_{\varepsilon,1}\| + \|\hat{v} - v^*_{\varepsilon',1}\|)^2 \geq \frac{1}{4}\|v^*_{\varepsilon,1} - v^*_{\varepsilon',-1}\|^2 \geq (\alpha(M_0, \gamma_{\max}) - 1) \cdot \delta^2.$$

By Le Cam's lemma, we once again have

$$\inf_{\hat{v}_n \in \hat{\mathcal{V}}_{\mathbb{X}}} \sup_{\substack{(L,b) \in \mathbb{C}_{\text{approx}} \\ \mathbb{P}_{L,b} \in \mathbf{G}_{\text{var}}(\sigma_L, \sigma_b)}} \mathbb{E}\|\hat{v}_n - v^*\|^2 \geq (\alpha(M_0, \gamma_{\max}) - 1) \cdot \delta^2 \cdot (1 - d_{\text{TV}}(\mathbb{P}_{-1}^{(n)}, \mathbb{P}_1^{(n)})) \geq (1-\omega) \cdot (\alpha(M_0, \gamma_{\max}) - 1) \cdot \delta^2.$$

Putting together the two cases completes the proof.

6. Discussion

In this paper, we studied methods for computing approximate solutions to fixed-point equations in Hilbert spaces using methods that search over low-dimensional subspaces of the Hilbert space and that operate on stochastic observations of the problem data. We analyzed a standard stochastic approximation scheme involving Polyak–Ruppert averaging and proved nonasymptotic instance-dependent upper bounds on its mean-squared error. This upper bound involved a pure approximation error term, reflecting the discrepancy induced by searching over a finite-dimensional subspace as opposed to the Hilbert space, and an estimation error term, induced by the noisiness in the observations. We complemented this upper bound with an information-theoretic analysis that established instance-dependent lower bounds for both the approximation error and the estimation error. A noteworthy consequence of our analysis is that the optimal approximation factor in the oracle inequality is neither unity nor constant but a quantity depending on the projected population-level operator. By applying our general theorems, we showed oracle inequalities for three specific examples in statistical estimation: linear regression on a linear subspace, Galerkin methods for elliptic PDEs, and value function estimation via temporal difference methods in Markov reward processes.

The results of this paper leave open a number of directions for future work:

• This paper focused on the case of independently drawn observations. Another observation model, one that arises naturally in the context of reinforcement learning, is the Markov observation model. As discussed in Section 2.2.3, consider the problem with $L = \gamma P$ and $b = r$, where P is a Markov transition kernel, γ is the discount factor, and r is the reward function. The observed states and rewards in this setup are given by a single trajectory of the Markov chain P, as opposed to being drawn i.i.d. from the stationary distribution. It is known (Tsitsiklis and Van Roy [60]) that the resolvent formalism (a.k.a. TD(λ)) leads to an improved approximation factor with larger $\lambda \in [0,1)$. On the other hand, larger choices of λ may lead to larger variance and slower convergence for the stochastic approximation estimator, and a model selection problem exists (see section 2.2 in Szepesvári [59] for a detailed discussion). It is important for future work to extend our fine-grained risk bounds to the case of TD(λ) methods with Markov data. Leveraging the instance-dependent upper and lower bounds, one can also design and analyze estimators that achieve the optimal trade-off.

• This paper focused purely on oracle inequalities defined with respect to a subspace. However, the framework of oracle inequalities is far more general; in the context of statistical estimation, one can prove oracle inequalities for any star-shaped set with bounds on its metric entropy. (See section 13.3 in the monograph of Wainwright [65] for the general mechanism and examples.) For all three examples considered in Section 2.2, one might imagine

approximating solutions using sets with nonlinear structure, such as those defined by ℓ_1-constraints, Sobolev ellipses, or the function class representable by a given family neural networks. An interesting direction for future work is to understand the complexity of projected fixed-point equations defined by such approximating classes.

Acknowledgements

The authors thank Peter Bartlett for helpful discussions.

Endnotes

[1] Note that one can achieve an approximation factor arbitrarily close to 1 provided that $n \gg D$. One way to do so is as follows: form the plug-in estimate that solves the original fixed-point relation (1) on the sample averages $\frac{1}{n}\sum_{i=1}^n L_i$ and $\frac{1}{n}\sum_{i=1}^n b_i$, and then project this solution onto the subspace $\mathbb{S}$. In this paper, our principal interest—driven by the practical examples of Galerkin approximation and temporal difference learning—is in the regime $d \ll n \ll D$.

[2] As noted by Bradtke and Barto [12], this method can be understood as an instrumental variable method (Wooldridge [69]), and our results also apply to this more general setting.

[3] It should be noted that Galerkin methods apply to a broader class of problems, including linear PDEs of parabolic and hyperbolic type (Larsson and Thomée [34]), as well as kernel integral equations (Polydorides et al. [48, 49]).

[4] As a side remark, we note that our noise conditions can be further weakened, if desired, via a minibatching trick. To be precise, given any problem instance $\mathbb{P}_{L,b} \in \mathbf{G}_{\mathsf{var}}(\sigma_L, \sigma_b)$ and any integer $m > 0$, one could treat the sample mean of m independent samples as a single sample, resulting in a problem instance in the class $\mathbf{G}_{\mathsf{var}}\left(\frac{\sigma_L}{\sqrt{m}}, \frac{\sigma_b}{\sqrt{m}}\right)$. The same lower bound still applies to the class $\mathbf{G}_{\mathsf{var}}\left(\frac{\sigma_L}{\sqrt{m}}, \frac{\sigma_b}{\sqrt{m}}\right)$, at a cost of stronger dimension requirement $D \geq d + \frac{12}{\omega} n^2 m^2$.

[5] Note that the stochastic approximation iterates are invariant under translation, and consequently, we can assume, without loss of generality, that $\overline{v} = 0$.

[6] In the typical application of finite-element methods, basis functions based on local interpolation are widely used (Brenner and Scott [13]). These basis functions can have large sup-norm, but via application of the Walsh–Hadamard transform, a new basis can be obtained satisfying Condition (40) with dimension-independent constants. Because the stochastic approximation algorithm is invariant under orthogonal transformation, this modification is only for the convenience of analysis and does not change the algorithm itself.

[7] Because the functions ψ_i are linearly independent, we have $B > 0$.

References

[1] Arridge S, Maass P, Öktem O, Schönlieb C-B (2019) Solving inverse problems using data-driven models. *Acta Numer.* 28(May):1–174.
[2] Bartlett PL, Bousquet O, Mendelson S (2005) Local Rademacher complexities. *Ann. Statist.* 33(4):1497–1537.
[3] Bellec PC (2018) Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist.* 46(2):745–780.
[4] Benveniste A, Métivier M, Priouret P (2012) *Adaptive Algorithms and Stochastic Approximations*, translated by Wilson SS (Springer Science & Business Media, New York).
[5] Bertsekas DP (2011) Temporal difference methods for general projected equations. *IEEE Trans. Automatic Control* 56(9):2128–2139.
[6] Bertsekas DP (2018) Proximal algorithms and temporal difference methods for solving fixed point problems. *Comput. Optim. Appl.* 70(3):709–736.
[7] Bertsekas DP (2019) *Reinforcement Learning and Optimal Control* (Athena Scientific, Belmont, MA).
[8] Bhandari J, Russo D, Singal R (2021) A finite time analysis of temporal difference learning with linear function approximation. *Oper. Res.* 69(3):950–973.
[9] Borkar VS (2009) *Stochastic Approximation: A Dynamical Systems Viewpoint* (Hindustan Book Agency, New Delhi, India).
[10] Bousquet O, Kane D, Moran S (2019) The optimal approximation factor in density estimation. Beygelzimer A, Hsu D, eds. *Proc. Machine Learn. Res.*, vol. 99, Conference on Learning Theory, Phoenix, AZ (PMLR), 318–341.
[11] Boyan JA (2002) Technical update: Least-squares temporal difference learning. *Machine Learn.* 49(2–3):233–246.
[12] Bradtke SJ, Barto AG (1996) Linear least-squares algorithms for temporal difference learning. *Machine Learn.* 22(1–3):33–57.
[13] Brenner SC, Scott LR (2007) *The Mathematical Theory of Finite Element Methods*, 3rd ed. (Springer Science & Business Media, New York).
[14] Bunea F, Tsybakov AB, Wegkamp MH (2007) Aggregation for Gaussian regression. *Ann. Statist.* 35(4):1674–1697.
[15] Bunea F, Tsybakov A, Wegkamp M (2007) Sparsity oracle inequalities for the Lasso. *Electronic J. Statist.* 1:169–194.
[16] Céa J (1964) Approximation variationnelle des problèmes aux limites. (In French.) *Ann. Inst. Fourier* 14(2):345–444.
[17] Chan SO, Diakonikolas I, Servedio RA, Sun X (2014) Near-optimal density estimation in near-linear time using variable-width histograms. Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger KQ, eds. *Advances in Neural Information Processing Systems*, Vol. 27 (Curran Associates, Red Hook, NY), 1844–1852.
[18] Dalalyan AS, Salmon J (2012) Sharp oracle inequalities for aggregation of affine estimators. *Ann. Statist.* 40(4):2327–2355.
[19] Dalalyan AS, Sebbar M (2018) Optimal Kullback–Leibler aggregation in mixture density estimation by maximum likelihood. *Math. Statist. Learn.* 1(1):1–35.
[20] Darolles S, Fan Y, Florens J-P, Renault E (2011) Nonparametric instrumental regression. *Econometrica* 79(5):1541–1565.
[21] Duchi JC, Ruan F (2021) Asymptotic optimality in stochastic optimization. *Ann. Stat.* 49(1):21–48.
[22] Fletcher CAJ (1984) Computational Galerkin methods. *Computational Galerkin Methods* (Springer, New York), 72–85.
[23] Galerkin BG (1915) Series solution of some problems of elastic equilibrium of rods and plates. (In Russian.) *Vestnik Inzhenerov i Tekhnikov* 19(7):897–908.

[24] Giordano M, Nickl R (2020) Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inverse Problems* 36(8):085001.
[25] Kaltenbacher B, Kirchner A, Vexler B (2011) Adaptive discretizations for the choice of a Tikhonov regularization parameter in nonlinear inverse problems. *Inverse Problems* 27(12):125008.
[26] Khamaru K, Pananjady A, Ruan F, Wainwright MJ, Jordan MI (2021) Is temporal difference learning optimal? An instance-dependent analysis. *SIAM J. Math. Data Sci.* 3(4):1013–1040.
[27] Klopp O, Tsybakov AB, Verzelen N (2017) Oracle inequalities for network models and sparse graphon estimation. *Ann. Statist.* 45(1):316–354.
[28] Koltchinskii V (2006) Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.* 34(6):2593–2656.
[29] Koltchinskii V (2011) *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole d'Eté de Probabilités de Saint-Flour XXXVIII-2008* (Springer, Berlin).
[30] Konda VR, Tsitsiklis JN (2000) Actor-critic algorithms. Solla S, Leen T, Müller K, eds. *Advances in Neural Information Processing Systems*, Vol. 12 (MIT Press, Cambridge, MA), 1008–1014.
[31] Krasnosel'skii MA, Vaĭnikko GM, Zabreĭko RP, Ruticki YB, Stet'senko VY (1972) *Approximate Solution of Operator Equations*, translated by Louvish D (Wolters-Noordhoff Publishing, Groningen, Netherlands).
[32] Lai TL (2003) Stochastic approximation. *Ann. Statist.* 31(2):391–406.
[33] Lakshminarayanan C, Szepesvári C (2018) Linear stochastic approximation: How far does constant step-size and iterate averaging go? Storkey A, Perez-Cruz F, eds. *Internat. Conf. Artificial Intelligence Statist.*, Playa Blanca, Lanzarote, Canary Islands (PMLR), 1347–1355.
[34] Larsson S, Thomée V (2008) *Partial Differential Equations with Numerical Methods* (Springer Science & Business Media, New York).
[35] Li S, Cai T, Li H (2021) Transfer learning for high-dimensional linear regression: Prediction, estimation, and minimax optimality. *J. Roy. Statist. Soc. Series B, Statist. Methodol.* 84(1):149–173.
[36] Li CJ, Mou W, Wainwright MJ, Jordan MI (2022) ROOT-SGD: Sharp nonasymptotics and asymptotic efficiency in a single algorithm. *Proc. 35th Conf. Learn. Theory*, Proceedings of Machine Learning Research Series (PMLR, London), 909–981.
[37] Lung R, Wu Y, Kamilis D, Polydorides N (2020) A sketched finite element method for elliptic models. *Comput. Methods Appl. Mech. Engrg.* 364(June):112933.
[38] Massart P (2007) *Concentration Inequalities and Model Selection* (Springer, Berlin).
[39] Massart P, Nédélec É (2006) Risk bounds for statistical learning. *Ann. Statist.* 34(5):2326–2366.
[40] Mou W, Li CJ, Wainwright MJ, Bartlett PL, Jordan MI (2020) On linear stochastic approximation: Fine-grained Polyak-Ruppert and non-asymptotic concentration. Abernethy J, Agarwal S, eds. *Proc. 33rd Conf. Learn. Theory* (PMLR), 2947–2997.
[41] Moulines É, Bach FR (2011) Non-asymptotic analysis of stochastic approximation algorithms for machine learning. Shawe-Taylor J, Zemel R, Bartlett P, Pereira F, Weinberger KQ, eds. *Advances in Neural Information Processing Systems*, Vol. 24 (Curran Associates, Red Hook, NY), 451–459.
[42] Munos R, Szepesvári C (2008) Finite-time bounds for fitted value iteration. *J. Machine Learn. Res.* 9(May):815–857.
[43] Nemirovski A, Juditsky A, Lan G, Shapiro A (2009) Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.* 19(4):1574–1609.
[44] Nickl R (2017) On Bayesian inference for some statistical inverse problems with partial differential equations. *Bernoulli News* 24(2):5–9.
[45] Pananjady A, Wainwright MJ (2021) Instance-dependent ℓ_∞-bounds for policy evaluation in tabular reinforcement learning. *IEEE Trans. Inform. Theory* 67(1):566–585.
[46] Polyak BT (1990) A new method of stochastic approximation type. (In Russian.) *Automatika i Telemekh* 51(7):98–107.
[47] Polyak BT, Juditsky AB (1992) Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* 30(4):838–855.
[48] Polydorides N, Wang M, Bertsekas DP (2009) Approximate solution of large-scale linear inverse problems with Monte Carlo simulation. LIDS Report 2822, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA.
[49] Polydorides N, Wang M, Bertsekas DP (2012) A quasi Monte Carlo method for large-scale inverse problems. *Monte Carlo and Quasi-Monte Carlo Methods 2010* (Springer,), 623–637.
[50] Rakhlin A, Sridharan K, Tsybakov AB (2017) Empirical entropy, minimax regret and minimax risk. *Bernoulli* 23(2):789–824.
[51] Rigollet P, Hütter J-C (2015) 18.S997: High dimensional statistics. Lecture notes (spring 2015), Massachusetts Institute of Technology, Cambridge, MA.
[52] Robbins H, Sutton M (1951) A stochastic approximation method. *Ann. Math. Statist.* 22(3):400–407.
[53] Rummery GA, Niranjan M (1994) On-line Q-learning using connectionist systems. Technical Report CUED/F-INFENG/TR 166, Cambridge University Department of Engineering, University of Cambridge, Cambridge, UK.
[54] Ruppert D (1988) Efficient estimations from a slowly convergent Robbins-Monro process. Technical Report 781, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY.
[55] Scherrer B (2010) Should one compute the temporal difference fix point or minimize the Bellman residual? The unified oblique projection view. Fürnkranz J, Joachims T, eds. *Proc. 27th Internat. Conf. Internat. Conf. Machine Learn.* (Omnipress, Madison, WI), 959–966.
[56] Srikant R, Ying L (2019) Finite-time error bounds for linear stochastic approximation and TD learning. *32nd Annual Conf. Learn. Theory*, Phoenix, AZ (PMLR), 2803–2830.
[57] Sutton RS (1988) Learning to predict by the methods of temporal differences. *Machine Learn.* 3(1):9–44.
[58] Sutton RS, Maei HR, Precup D, Bhatnagar S, Silver D, Szepesvári C, Wiewiora E (2009) Fast gradient-descent methods for temporal-difference learning with linear function approximation. Bottou L, Littman M, eds. *Proc. 26th Annual Internat. Conf. Machine Learn.* (ACM, New York), 993–1000.
[59] Szepesvári C (2010) *Algorithms for Reinforcement Learning* (Morgan & Claypool Publishers, San Rafael, CA).
[60] Tsitsiklis JN, Van Roy B (1997) Analysis of temporal-diffference learning with function approximation. Jordan MI, Petsche T, eds. *Proc. 9th Internat. Conf.* Neural Inform. Processing Systems (MIT Press, Cambridge, MA), 1075–1081.
[61] Tsitsiklis JN, Van Roy B (1999) Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives. *IEEE Trans. Automatic Control* 44(10):1840–1851.
[62] Tsybakov AB (2004) Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* 32(1):135–166.
[63] van der Vaart AW (2000) *Asymptotic Statistics* (Cambridge University Press, Cambridge, UK).

[64] Van Roy B (2006) Performance loss bounds for approximate value iteration with state aggregation. *Math. Oper. Res.* 31(2):234–244.
[65] Wainwright MJ (2019) *High-Dimensional Statistics: A Non-Asymptotic Viewpoint* (Cambridge University Press, Cambridge, UK).
[66] Wainwright MJ (2019) Stochastic approximation with cone-contractive operators: Sharper ℓ_∞-bounds for Q-learning. Preprint, submitted May 15, https://doi.org/10.48550/arXiv.1905.06265.
[67] Wainwright MJ (2019) Variance-reduced Q-learning is minimax optimal. Preprint, submitted June 11, https://doi.org/10.48550/arXiv.1906.04697.
[68] Watkins CJCH, Dayan P (1992) Q-Learning. *Machine Learn.* 8(3–4):279–292.
[69] Wooldridge JM (2016) *Introductory Econometrics: A Modern Approach*. 7th ed. (Nelson Education, Scarborough, ON, Canada).
[70] Yatracos YG (1985) Rates of convergence of minimum distance estimators and Kolmogorov's entropy. *Ann. Statist.* 13(2):768–774.
[71] Yu H, Bertsekas DP (2010) Error bounds for approximations from projected linear equations. *Math. Oper. Res.* 35(2):306–329.
[72] Zhu B, Jiao J, Tse D (2020) Deconstructing generative adversarial networks. *IEEE Trans. Inform. Theory* 66(11):7155–7179.