A unified framework for temporal difference methods

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/ADPRL.2009.4927518">http://dx.doi.org/10.1109/ADPRL.2009.4927518</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Thu Dec 20 20:03:33 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/58831">http://hdl.handle.net/1721.1/58831</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>


A Unified Framework for Temporal Difference Methods

Dimitri P. Bertsekas
Laboratory for Information and Decision Systems (LIDS)
Massachusetts Institute of Technology
MA 02139, USA
Email: dimitrib@mit.edu

Abstract—We propose a unified framework for a broad class of methods to solve projected equations that approximate the solution of a high-dimensional fixed point problem within a subspace \( S \) spanned by a small number of basis functions or features. These methods originated in approximate dynamic programming (DP), where they are collectively known as temporal difference (TD) methods. Our framework is based on a connection with projection methods for monotone variational inequalities, which involve alternative representations of the subspace \( S \) (feature scaling). Our methods admit simulation-based implementations, and even when specialized to DP problems, include extensions/new versions of the standard TD algorithms, which offer some special implementation advantages and reduced overhead.

I. INTRODUCTION

We consider the approximation of a fixed point of a mapping \( T : \mathbb{R}^n \mapsto \mathbb{R}^n \) by solving the projected equation

\[ x = \Pi T(x), \]

where \( \Pi \) denotes projection onto a closed subset of a subspace \( S \). The projection is with respect to a weighted Euclidean norm \( \| \cdot \|_\xi \), where \( \xi \) is a probability distribution vector with positive components (i.e., \( \sum_{i=1}^n \xi_i x_i^2 \)). We will assume throughout that \( \Pi T \) is a contraction with respect to some norm, so the projected equation \( x = \Pi T(x) \) has a unique solution.

The projected equation approach is common in approximate DP, where \( T \) is a (possibly multistep) DP/Bellman operator associated with a fixed policy of a Markovian decision problem (MDP), and \( x \) is the cost vector of that policy [see the books by Bertsekas and Tsitsiklis [BeT96], and Sutton and Barto [SuB98]; Bertsekas [Ber07] provides a recent textbook treatment and up-to-date references]. There are some well-known algorithms for solving the corresponding projected equation [such as TD(\( \lambda \)) (proposed by Sutton [Sut88]), least squares policy evaluation (LSPE; originally proposed by Bertsekas and Ioffe [BeI96], and followed up by Nedić and Bertsekas, [NeBo3], Bertsekas, Borkar, and Nedić [BBN04], and Yu and Bertsekas [YuBo06]), least squares temporal differences (LSTD; originally proposed by Bradtke and Barto [BrB96], and followed up by Boyan [Boy02], and Nedić and Bertsekas [NeBo3]), and the Fixed Point Kalman Filter (FPKF; proposed by Choi and Van Roy [ChV06])]. The first three of these algorithms have been recently extended for approximate solution of general linear (and some nonlinear) fixed point problems (see Bertsekas and Yu [BeY08]).

Earlier work has assumed a fixed representation of the subspace \( S \):

\[ S = \{ \Phi r \mid r \in \mathbb{R}^s \}, \tag{1} \]

where \( \Phi \) is an \( n \times s \) matrix of full rank whose columns are viewed as basis functions or features. One of the purposes of the present paper is to explore the effect of alternative representations

\[ S = \{ \Psi v \mid v \in \mathbb{R}^s \}, \tag{2} \]

where \( \Phi = \Psi B \). Here \( \Psi \) is an \( n \times s \) matrix and \( B \) is an \( s \times s \) matrix with range equal to \( \mathbb{R}^s \). It can be seen that the range spaces of \( \Phi \) and \( \Psi \) coincide, so \( S \) can be alternatively represented as in Eq. (2). For a proof, note that any vector of the form \( \Phi r \) can be expressed as \( \Psi v \), where \( \Phi = \Psi B \). Here \( \Psi \) is an \( n \times s \) matrix and \( B \) is an \( s \times s \) matrix with range equal to \( \mathbb{R}^s \). It can be seen that the range spaces of \( \Phi \) and \( \Psi \) coincide, so \( S \) can be alternatively represented as in Eq. (2).

As a vehicle for analysis and algorithmic development, we introduce in Section 2 a connection between projected equations and a special type of monotone variational inequalities (VI for short). We then consider some known scaled projection methods for this type of VI, which are particularly relevant to subspace approximation. This leads to some interesting new methods developed in Sections 3 and 4. These methods include scaled versions of the projected Jacobi/value iteration method of approximate DP and its corresponding simulation-based version (LSPE), as well as their extensions to more general fixed point problems. A possible benefit from the scaled versions is that they need not involve a potentially costly matrix inversion at each iteration.

In Section 3 we specialize the deterministic projection algorithms of Section 2 to linear fixed point problems,
including policy evaluation problems in approximate DP. The deterministic algorithms of Section 3 are somewhat impractical for large dimension $n$, as they require $n$-dimensional linear algebra calculations, so in Section 4 we develop simulation-based versions, which require low ($s$-dimensional) calculations only. In the process we recover the existing TD methods for approximate DP and we explore the effect of feature scaling on them. We show that the LSPE-type projected Jacobi method is scale-free, but the performance and the range of stepsize choices of the other methods are greatly affected by feature scaling. When simulation-based implementations are introduced, the slower speed of simulation dominates, and the manner in which scaling affects performance is not well understood at present. We briefly mention also the connection with the TD and FPKF algorithms, but note that an example given in Yu and Bertsekas [YuB07] shows that even with optimal scaling, TD and FPKF have worse convergence rate than LSPE, which together with LSTD, has optimal convergence rate.

A byproduct of our analysis is that the common assumption that $\Phi$ has full rank $s$ is unnecessary for the convergence of TD-type methods. It is necessary only to ensure that inverses appearing in various implementations are well-defined.

II. PROJECTED EQUATIONS AND RELATION TO VARIATIONAL INEQUALITIES

The standard VI problem is to find a vector $r^* \in R$ such that

$$F(r^*)' (r - r^*) \geq 0, \quad \forall r \in R,$$

where $R$ is a closed convex set and $F : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a given function (in our notation $\mathbb{R}$ is the $s$-dimensional Euclidean space, all vectors in $\mathbb{R}$ are viewed as column vectors and a prime denotes transposition). A major class of methods for solving the VI (3) is projection methods of the form

$$r_{k+1} = P_{D,R} [r_k - \gamma D^{-1} F(r_k)],$$

where $\gamma$ is a positive stepsize, $D$ is a positive definite symmetric matrix, and $P_{D,R} [\cdot]$ denotes projection on $X$ with respect to the norm $\| r \|_D = \sqrt{r^T D r}$. The standard convergence result for projection methods (see e.g., [BeT89], Section 3.5.3, or [PaF03]) assumes that $F$ is Lipschitz continuous and strongly monotone, in the sense that for some scalars $L > 0$ and $\beta > 0$, we have

$$\| f(r_1) - f(r_2) \| \leq L \| r_1 - r_2 \|, \quad \forall r_1, r_2 \in R,$$

$$(f(r_1) - f(r_2))' (r_1 - r_2) \geq \beta \| r_1 - r_2 \|^2, \quad \forall r_1, r_2 \in R,$$

(here $\| \cdot \|$ is some norm, e.g., the standard Euclidean norm). Then it can be shown that the VI (3) has a unique solution $r^*$, and there exists $\bar{\gamma} > 0$ such that $r_k \to r^*$ linearly for each constant stepsize $\gamma$ in the range $(0, \bar{\gamma})$ [i.e., $\{\| r_k - r^* \| \}$ converges to 0 at least as fast as a geometric progression]. The strong monotonicity assumption is essential for this - just monotonicity (i.e., $\beta = 0$) may result in divergence (see e.g., [BeT89], p. 270).

In an interesting specially structured class of VIs, $F$ has the form

$$F(r) = \Phi f(\Phi r),$$

where $\Phi$ is an $n \times m$ matrix, and $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ is Lipschitz continuous and strongly monotone. Such a VI may result from a VI of the form

$$f(x^*)' (x - x^*) \geq 0, \quad \forall x \in X,$$

via a transformation of variables $x = \Phi r$; see Bertsekas and Gafni [BeG82], where this special type of VI and corresponding projection methods were first studied.

Note that if $f$ is Lipschitz continuous, so is $F$, but even if $f$ is strongly monotone, $F$ need not be, and the solution of the corresponding VI may not be unique. Despite this fact, the corresponding projection method,

$$r_{k+1} = P_{D,R} [r_k - \gamma D^{-1} \Phi f(\Phi r_k)],$$

has been shown in [BeG82] to have the same type of convergence property, i.e., that there exists $\tilde{\gamma} > 0$ such that $r_k \to r^*$ linearly for each $\gamma \in (0, \tilde{\gamma}]$, where $r^*$ is some solution of

$$f(\Phi r^*)' (\Phi r - r^*) \geq 0, \quad \forall r \in R,$$

provided $f$ is strongly monotone and $R$ is a polyhedral set (the polyhedral assumption is essential).

We now consider the projected equation

$$x = \Pi T(x)$$

and cast it as a variational inequality. Here for additional generality, we assume that $\Pi$ is the projection operation (with respect to the norm $\| \cdot \|_\xi$) onto a subset $\hat{S} \subset S$ of the form

$$\hat{S} = \{ \Phi r \mid r \in R \},$$

where $R$ is a polyhedral subset of $\mathbb{R}$. We assume that $\Pi T$ is a contraction mapping with respect to some norm, so the projected equation has a unique solution $\Phi r^* \in \hat{S}$ (even though $r^*$ may not be unique if $\Phi$ does not have full rank).

Note that some of the error bounds given in [BeY08] for the case where $\hat{S} = S$ apply to the more general case considered here. In particular, to evaluate the distance between $\Phi r^*$ and a fixed point $x^*$ of $T$, assuming that $\Pi T$ is a contraction with respect to a norm $\| \cdot \|$ with modulus $\alpha \in [0, 1)$, we have

$$\| x^* - \Phi r^* \| \leq \| x^* - \Pi x^* \| + \| \Pi T(x^*) - \Pi T(\Phi r^*) \|$$

$$\leq \| x^* - \Pi x^* \| + \alpha \| x^* - \Phi r^* \|,$$

so that

$$\| x^* - \Phi r^* \| \leq \frac{1}{1 - \alpha} \| x^* - \Pi x^* \|.$$
By the properties of projection, \( x^* \) satisfies \( x^* = \Pi T(x^*) \) if and only if \( x^* \in S \) and the vector \( x^* - T(x^*) \) forms a nonnegative inner product with all vectors \( x - x^* \) with \( x \in S \), i.e.,

\[
(x^* - T(x^*))^\top \Xi(x - x^*) \geq 0, \quad \forall \ x \in \hat{S},
\]

where \( \Xi \) is the diagonal matrix that has the components of \( \xi \) along the diagonal (the projection norm here is \( \| \cdot \|_\xi \) and the associated inner product of two vectors \( x_1, x_2 \) is \( x_1^\top \Xi x_2 \)). This is a VI, which in view of the form of the set \( \hat{S} \), can be written in the special form of Eq. (6): finding \( r^* \in R \) such that

\[
f(\Phi r^*) \Phi (r - r^*) \geq 0, \quad \forall \ r \in R,
\]

where

\[
f(x) = \Xi(x - T(x)).
\]

The projection algorithm of Eq. (5) takes the form

\[
r_{k+1} = P_{D,R} \left[ r_k - \gamma D^{-1} \Phi \Xi (\Phi r_k - T(\Phi r_k)) \right].
\]

We now show that contraction properties of \( T \) or \( \Pi T \) imply that \( f \) is strongly monotone over \( \hat{S} \), which is what is required for the scaled projection algorithm (10) to be convergent in the sense discussed earlier.

**Proposition 1.** Assume that \( T \) is a contraction with respect to the norm \( \| \cdot \|_\xi \). Then the function \( f \) of Eq. (9) is strongly monotone over the set \( \hat{S} \).

**Proof:** Let \( \alpha \in [0,1) \) be the modulus of contraction of \( T \). For any two vectors \( x_1, x_2 \in \hat{S} \),

\[
(f(x_1) - f(x_2))^\top (x_1 - x_2) = (x_1 - T(x_1) - x_2 + T(x_2))^\top \Xi(x_1 - x_2)
\]

\[
= (x_1 - x_2)^\top \Xi(x_1 - x_2) - (T(x_1) - T(x_2))^\top \Xi(x_1 - x_2)
\]

\[
\geq \| x_1 - x_2 \|_\xi^2 - \| T(x_1) - T(x_2) \|_\xi \| x_1 - x_2 \|_\xi
\]

\[
\geq \| x_1 - x_2 \|_\xi^2 - \alpha \| x_1 - x_2 \|_\xi^2
\]

\[
= (1 - \alpha) \| x_1 - x_2 \|_\xi^2,
\]

where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality follows from the contraction property of \( T \). Since \( \alpha \in [0,1) \), this shows that \( f \) is strongly monotone on \( \hat{S} \).

In the special case where \( T \) is the entire space \( \mathbb{R}^n \), i.e., \( \hat{S} = S \) and \( \Pi \) is projection on the subspace \( S \), it is sufficient that \( \Pi T \) rather than \( T \) be a contraction. The origin of the following proposition can be traced to the convergence proof of TD(\( \lambda \)) by Tsitsiklis and Van Roy [TsV97] (Lemma 9); see also [BeY08], Prop. 5.

**Proposition 2.** Assume that \( \hat{S} = S \) and that \( \Pi T \) is a contraction with respect to the norm \( \| \cdot \|_\xi \). Then the function \( f \) of Eq. (9) is strongly monotone over the subspace \( S \).

**Proof:** Let \( \alpha \in [0,1) \) be the modulus of contraction of \( \Pi T \), and note that we have

\[
(T(x) - \Pi T(x))^\top \Xi x = 0, \quad \forall \ x, \tilde{x} \in S.
\]

since vectors of the form \( x - \Pi x \) are orthogonal (with respect to the norm \( \| \cdot \|_\xi \)) to vectors in \( S \). We use this equation as an intermediate step in the proof of the preceding proposition to obtain the desired conclusion.

We have for any two vectors \( x_1, x_2 \in S \),

\[
(f(x_1) - f(x_2))^\top (x_1 - x_2) = (x_1 - T(x_1) - x_2 + T(x_2))^\top \Xi(x_1 - x_2)
\]

\[
= (x_1 - x_2)^\top \Xi(x_1 - x_2) - (T(x_1) - T(x_2))^\top \Xi(x_1 - x_2)
\]

\[
\geq \| x_1 - x_2 \|_\xi^2 - \| T(x_1) - T(x_2) \|_\xi \| x_1 - x_2 \|_\xi
\]

\[
\geq \| x_1 - x_2 \|_\xi^2 - \alpha \| x_1 - x_2 \|_\xi^2
\]

\[
= (1 - \alpha) \| x_1 - x_2 \|_\xi^2,
\]

where the third equation follows from Eq. (11), the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality follows from the contraction property of \( \Pi T \). This shows that \( f \) is strongly monotone on \( S \).

There are well-known cases in approximate DP where \( \Pi T \) is a contraction with respect to \( \| \cdot \|_\xi \) (see [BeT96], [TsV97], [Ber07]). An example arises in discounted MDP, where \( T \) is of the form \( T(x) = \alpha P x + b \), with \( \alpha \in (0,1) \), \( P \) is a transition probability matrix of an ergodic Markov chain, and \( \xi \) is the steady-state distribution of that chain. Reference [BeY08] provides several general criteria for verifying that \( \Pi T \) is a contraction, beyond the MDP context.

**III. Projection Methods for Projected Equations**

While there are interesting problems where the set \( R \) is a strict subset of \( \mathbb{R}^n \) and/or \( T \) is a nonlinear contraction (e.g., optimal stopping problems; see [BeT96], Section 6.8, or [Ber07], Section 6.4), we will focus on the case where \( R = \mathbb{R}^n \) and \( T \) is linear of the form

\[
T(x) = Ax + b,
\]

where \( A \) is an \( n \times n \) matrix and \( b \) is a vector in \( \mathbb{R}^n \). The algorithms and analysis of this and the next section extend straightforwardly to the case where \( T \) is a linear contraction with respect to \( \| \cdot \|_\xi \) and \( R \) is a polyhedral strict subset of \( \mathbb{R}^n \).
We will assume for the remainder of the paper that \( \Pi T \) is a contraction with respect to \( \| \cdot \|_\epsilon \), so that Prop. 2 applies to the VI (8)-(9). Since there is no constraint set \( \mathcal{R} \), this VI becomes the linear equation \( \Phi f(\Phi r^*) = 0 \), which using the fact
\[
f(x) = \Xi(x - T(x)) = \Xi(x - Ax - b),
\]
can be written as
\[
\Phi' \Xi(I - A) \Phi r^* = \Phi' \Xi b.
\] (12)

Since the projection \( P_{D,R}[] \) is trivially the identity, the projection method (10) takes the form
\[
r_{k+1} = r_k + \gamma D^{-1} \Phi' \Xi (A - I) \Phi r_k + b),
\] (13)
and can also be written as an iteration in the high-dimensional vector \( x \in \mathbb{R}^n \):
\[
x_{k+1} = x_k + \gamma \Phi D^{-1} \Phi' \Xi (A - I) x_k + b),
\] (14)
where we denote \( x_k = \Phi r_k \). Based on the results of [BeO82] mentioned earlier and Prop. 2, there is a range of stepizes \( (0, \gamma) \) such that the sequence \( \{r_k\} \) converges to some solution of Eq. (12). If \( \Phi \) does not have full rank, the matrix \( \Phi' \Xi (I - A) \Phi \) is not invertible, and there will be multiple solutions. Otherwise, \( \{r_k\} \) converges to the unique solution \( r^* \), given by
\[
r^* = (\Phi' \Xi (I - A) \Phi)^{-1} \Phi' \Xi b.
\] (15)

Note that a full rank for the feature matrix \( \Phi \) is not essential for convergence in general, and does not appear to have a significant effect on the algorithm, unless it is required for the definition of \( D \).

A. Some Special Cases

Generally, the stepsize range for convergence is unknown and strongly depends on the matrix \( D \). This is because, for convergence, the iteration matrix
\[
I + \gamma D^{-1} \Phi' \Xi (A - I) \Phi
\]
should be a contraction with respect to some norm. However, in the special case where
\[
D = \Phi' \Xi \Phi
\]
(assuming \( \Phi \) has full rank so that \( \Phi' \Xi \Phi \) is invertible), the range of appropriate stepsizes includes the interval \((0, 1)\). The reason is that for this choice of \( D \), the iteration (13) is equivalent to the fixed point iteration
\[
x_{k+1} = (1 - \gamma)x_k + \gamma \Pi T(x_k),
\] (16)
where \( x_k = \Phi r_k \). To verify this, we may write the projection mapping explicitly as
\[
\Pi = \Phi (\Phi' \Xi \Phi)^{-1} \Phi' \Xi.
\]
and compare with iteration (13), which for \( D = \Phi' \Xi \Phi \), takes the form:
\[
r_{k+1} = r_k + \gamma (\Phi' \Xi \Phi)^{-1} \Phi' \Xi (A - I) \Phi r_k + b).
\] (17)
For \( \gamma = 1 \), this is the projected Jacobi method
\[
x_{k+1} = \Pi T(x_k),
\]
which was studied in [BeY08], together with its simulation-based analog, to be discussed in the next section. In the case of a discounted MDP, where \( A = \alpha P \), \( \alpha \in (0, 1) \) is a discount factor, and \( P \) is the transition matrix of the associated Markov chain (see [Ber07], [Put94]), the method reduces to the projected value iteration method, and is related to the LSPE method, which is its simulation-based implementation (see [Ber07] for a textbook account).

Another interesting special case of the projection iteration (13) is when \( D \) is the identity:
\[
r_{k+1} = r_k + \gamma \Phi' \Xi (A - I) \Phi r_k + b).
\] (18)
We refer to this as the simple projection iteration, and we note that it can be related to TD(\( \lambda \)) (see [BeY08]).

B. Effects of Feature Scaling

We now consider the effect of different basis function representations, i.e., feature scaling, on the projection iteration (13). To this end, we represent the subspace \( S \) with two different feature matrices \( \Phi \) and \( \Psi \), related by
\[
\Phi = \Psi B,
\]
[cf. Eqs. (1)-(2)]. We compare the corresponding high-dimensional sequences
\[
x_{k,\Phi} = \Phi r_k, \quad x_{k,\Psi} = \Psi v_k,
\]
where \( r_k \) and \( v_k \) are generated by corresponding projection iterations:
\[
r_{k+1} = r_k + \gamma D^{-1}_\Phi \Phi' \Xi ((A - I) \Phi r_k + b),
\]
and
\[
v_{k+1} = v_k + \gamma D^{-1}_\Psi \Psi' \Xi ((A - I) \Psi v_k + b).
\]
We have [cf., Eq. (14)]
\[
x_{k+1,\Phi} = x_{k,\Phi} + \gamma \Phi D^{-1}_\Phi \Phi' \Xi ((A - I) x_{k,\Phi} + b),
\] (19)
and
\[
x_{k+1,\Psi} = x_{k,\Psi} + \gamma \Psi D^{-1}_\Psi \Psi' \Xi ((A - I) x_{k,\Psi} + b).\] (20)
Using the equation \( \Phi = \Psi B \), Eq. (19) is written as
\[
x_{k+1,\Phi} = x_{k,\Phi} + \gamma \Psi B D^{-1}_\Phi B' \Phi' \Xi ((A - I) x_{k,\Phi} + b),
\]
which compared with Eq. (20), shows that if the scaling matrices satisfy
\[
D^{-1}_\Psi = B D^{-1}_\Phi B',
\] (21)
the two scaled projection iterations (19) and (20) produce identical results within the high-dimensional space \( x_{k,\Phi} = x_{k,\Psi} \) for all \( k \), assuming that \( x_0,\Phi = x_0,\Psi \). In conclusion, alternative choices of feature scaling
correspond to alternative choices of the scaling matrix $D$.

Another observation, is that given a feature matrix $\Phi$ that has full rank, the entire class of projection iterations can be derived from the simple projection iteration (18) by generating scaling matrices as
\[ D^{-1} = BB' \]
from square feature scaling matrices $B$ [cf. Eq. (21)]. If $\Phi$ does not have full rank and $\Psi$ does, $B$ cannot be chosen square. In this case the simple projection iteration (18) is equivalent to an iteration executed over a lower-dimensional feature space corresponding to $\Psi$ with $\Phi = \Psi B$ and $BB'$ equal to the identity. Indeed, it can be seen that the iterates $r_k$ generated by Eq. (18) lie on the affine set
\[ r_0 + \text{Range}(\Phi'), \]
which has the same dimension as $S$, regardless of the choice of $\Phi$.

Consider now the effect of feature scaling on the projected Jacobi method. For feature matrices $\Phi$ and $\Psi$ that have full rank, and are related by $\Phi = \Psi B$, where $B$ is square and invertible, the corresponding scaling matrices are
\[ D_\Phi = \Phi' \Xi \Phi, \quad D_\Psi = \Psi' \Xi \Psi, \]
and we have
\[ D_\Phi = B' \Psi' \Xi \Psi B = B' D_\Psi B. \]
By taking the inverse of both sides above, we see that the scaling criterion (21) is satisfied. It follows that the projected Jacobi iteration is scale-free, and produces the same high-dimensional iteration sequence $\{x_k\}$ regardless of the representation of the subspace $S$. This is consistent with the fact that the iteration depends only on the projection operation $\Pi$ [cf. Eq. (16)], which is independent of how the subspace $S$ is represented.

IV. SIMULATION-BASED VERSIONS OF Projection Methods

We now consider implementations of the projection methods of the preceding section by using simulation and low-dimensional calculations. The approach for generating simulation-based implementations of deterministic algorithms for projected equations has been described in detail in [BeY08], and much of the analysis given there applies to our present context. Consequently our discussion will be brief.

The idea of the simulation approach is to sample a row and a column of the matrix $A$ at each iteration, then use the $k+1$ samples collected up to iteration $k$ to form a matrix $C_k$ and a vector $d_k$ that approximate the matrix
\[ C = \Phi' \Xi (I - A) \Phi, \]
and the vector
\[ d = \Phi' \Xi b, \]
and that appear in the projection iteration (13). The $k$th iteration is then approximated by
\[ r_{k+1} = r_k + \gamma D_k^{-1} (C_k r_k + d_k), \]
where $D_k$ is chosen so that $D_k \rightarrow D$, and $D$ is a positive definite symmetric scaling matrix.

One way to obtain $C_k$ and $d_k$, given in [BeY08], is to generate a sequence of indices $\{i_0, i_1, \ldots\}$, and a sequence of transitions between indices $\{ (i_0, j_0), (i_1, j_1), \ldots \}$. Any probabilistic mechanism may be used for this, subject to the following two requirements:

- The sequence $\{i_0, i_1, \ldots\}$ is generated according to the distribution $\zeta$, which defines the projection norm $\| \cdot \|_\zeta$, in the sense that with probability 1,
\[ \lim_{k \rightarrow \infty} \frac{\sum_{t=0}^{k} \delta(i_t = i)}{k + 1} = \xi_i, \quad i = 1, \ldots, n, \]
where $\delta(\cdot)$ denotes the indicator function $[\delta(E) = 1$ if the event $E$ has occurred and $\delta(E) = 0$ otherwise].
- The sequence $\{ (i_0, j_0), (i_1, j_1), \ldots \}$ is generated according to a certain stochastic matrix $P$ with transition probabilities $p_{ij}$ which satisfy
\[ p_{ij} > 0 \quad \text{if} \quad a_{ij} \neq 0, \]
in the sense that with probability 1,
\[ \lim_{k \rightarrow \infty} \frac{\sum_{t=0}^{k} \delta(i_t = i, j_t = j)}{\sum_{t=0}^{k} \delta(i_t = i)} = p_{ij}, \quad i, j = 1, \ldots, n. \]
Then $C_k$ and $d_k$ are computed as
\[ C_k = \frac{1}{k + 1} \sum_{t=0}^{k} \phi(i_t) \left( \frac{a_{i_t j_t}}{p_{i_t j_t}} \phi(j_t) - \phi(i_t) \right)' \]
and
\[ d_k = \frac{1}{k + 1} \sum_{i=0}^{k} \phi(i_t)b_{i_t}, \]
where we denote by $\phi(i)'$ the $i$th row of $\Phi$. Note that $C_k$ and $d_k$ can be updated recursively as new samples $\phi(i_t)$, $a_{i_t j_t}$, and $b_{i_t}$ are generated. In particular, we have
\[ C_k = \frac{1}{k + 1} \tilde{C}_k, \quad d_k = \frac{1}{k + 1} \tilde{d}_k, \]
where $\tilde{C}_k$ and $\tilde{d}_k$ are updated by
\[ \tilde{C}_k = \tilde{C}_{k-1} + \phi(i_k) \left( \frac{a_{i_k j_k}}{p_{i_k j_k}} \phi(j_k) - \phi(i_k) \right)' \]
and
\[ \tilde{d}_k = \tilde{d}_{k-1} + \phi(i_k)b_{i_k}. \]
It can be shown using simple law of large numbers arguments that $C_k \rightarrow C$ and $d_k \rightarrow d$ with probability 1 (see [BeY08]). In what follows, we will assume that $C_k$ and $d_k$ are chosen by using the preceding sampling scheme.

Note that in some cases, prominently arising in approximate DP, the probability vector $\xi$ is the steady-state distribution of a Markov chain, so it is only implicitly known. In this case, the sampling of rows should be done by simulating the Markov chain, as is standard in the approximate DP literature.

The matrix $D_k$ may also be computed by simulation in cases where $D$ depends on $\Xi$. An example is the projected Jacobi method (17), where the matrix $D = \Phi^T \Xi \Phi$ may be approximated by

$$D_k = \frac{1}{k+1} \sum_{t=0}^{k} \phi(i_t)\phi(i_t)' ,$$

or by

$$D_k = \frac{1}{k+1} \left( \delta I + \sum_{t=0}^{k} \phi(i_t)\phi(i_t)' \right) ,$$

where $\delta I$ is a positive multiple of the identity (to ensure that $D_k$ is positive definite). This is the approximate Jacobi method of [BeY08], which in the case of an approximate DP/policy evaluation problem, reduces to the LSPE method. It appears that this is the only special approximate DP/policy evaluation problem, reduces to Eq. (25) or (26) and $\gamma$ is chosen so that this is the basis for the LSTD algorithm in approximate DP. Note that our framework allows an algorithm of this type for the case where $\mathcal{K}$ is a general linear mapping, but uses only the last simulation sample. Similar to the TD-like method given above, it is reminiscent of the simulation-based projection method (24), but uses only the last simulation sample.

The difference is that the latter iteration uses only the last simulation sample, whereas the former iteration uses all the preceding samples.

Let us also mention the algorithm of Choi and Van Roy [ChV06], which when extended to our more general setting, has the form

$$r_{k+1} = r_k + \frac{\gamma}{k+1} \sum_{t=0}^{k} \phi(i_t) \left( \left( \frac{a_{t\gamma}(j_t) - \phi(j_t)}{p_{i_t\gamma j_t}} \right)' r_k + b_{i_t \gamma} \right) .$$

The difference is that the latter iteration uses only the last simulation sample, whereas the former iteration uses all the preceding samples.

Another interesting possibility is to use a diagonal scaling matrix $D_k$, thereby simplifying the matrix inversion in Eq. (24). One possible choice is a diagonal approximation to $\Phi^T \Xi \Phi$, obtained by discarding the off-diagonal terms of the matrix (25) or (26).

The convergence rate bottleneck of all the methods of the form (24) is the simulation, that is, the rate at which $C_k$ converges to $C$ and $d_k$ converges to $d$. This is true regardless of the choice of the limit $D$ of the scaling matrix $D_k$, as long as $\gamma$ is chosen so that the corresponding deterministic projection method (13) converges. The reason is that the corresponding deterministic method has a linear convergence rate, which is fast relative to the slow convergence of $C_k$ and $d_k$. Is there a choice of $D$ and $\gamma$ that is optimal from the point of view of rate of convergence? In the special case of a discounted or average cost MDP, there is one: it is the projected Jacobi method, where $D_k$ is given by Eq. (25) or (26) and $\gamma = 1$. This was conjectured in [BBN04] for the approximate DP/LSPE case, and was proved in [YuB06], based on results of Konda [Kon02] on the optimal convergence rate of LSTD, and a two-time scale argument (see, e.g., Borkar [Bor08], Chapter 6). Whether this optimality property of the projected Jacobi method holds more generally is an interesting subject for future research.

The special case where $D_k$ is the identity

$$r_{k+1} = r_k + \frac{\gamma}{k+1} \sum_{t=0}^{k} \phi(i_t) \left( \left( \frac{a_{t\gamma}(j_t) - \phi(j_t)}{p_{i_t\gamma j_t}} \right)' r_k + b_{i_t \gamma} \right) .$$

is reminiscent of the TD method of approximate DP, which when generalized for general linear equations has the form [BeY08]

$$r_{k+1} = r_k + \frac{\gamma}{k+1} \sum_{t=0}^{k} \phi(i_t) \left( \left( \frac{a_{t\gamma}(j_t) - \phi(j_t)}{p_{i_t\gamma j_t}} \right)' r_k + b_{i_t \gamma} \right) .$$

The difference is that the latter iteration uses only the last simulation sample, whereas the former iteration uses all the preceding samples.

Let us also mention the algorithm of Choi and Van Roy [ChV06], which when extended to our more general setting, has the form

$$r_{k+1} = r_k + \frac{\gamma}{k+1} \sum_{t=0}^{k} \phi(i_t) \left( \left( \frac{a_{t\gamma}(j_t) - \phi(j_t)}{p_{i_t\gamma j_t}} \right)' r_k + b_{i_t \gamma} \right) .$$

where $D_k$ is a positive definite symmetric matrix, which may be generated by Eqs. (25) or (26). Similar to the TD-like method given above, it is reminiscent of the simulation-based projection method (24), but uses only the last simulation sample.

We finally note the possibility of a simulation-based noniterative approach to solve the projected equation (12). Since this equation can be written as

$$C r^* = d,$$

where $C$ and $d$ are given by Eqs. (22) and (23), we may generate by simulation the matrix $C_k$ and vector $d_k$ as above, and approximate the solution $r^*$ by $C_k^{-1} d_k$ (assuming $C_k$ is invertible). This is the basis for the LSTD algorithm in approximate DP. Note that our framework allows an algorithm of this type for the case where $T$ is a general linear mapping, II{T} is a contraction with respect to $\| \cdot \|_{\xi}$, and the projection is onto a strict polyhedral subset $R$ of $S$: we simply estimate by simulation the matrix

$$C = \Phi^T (A - I) \Phi$$

and vector

$$d = \Phi^T \Xi b,$$
which appear in the corresponding VI (8)-(9),
\[(\Phi'\Xi (A-I)\Phi r^* - \Phi'\Xi b)' (r-r^*) \geq 0, \quad \forall r \in R,\]
and solve the (approximate) VI thus obtained.

V. CONCLUSIONS

In this paper we have considered the solution of projected equations that approximate large-scale fixed point problems. We have proposed a unifying framework for a broad class of methods that use simulation and low-order calculations. Our framework is based on a connection with variational inequalities, which lends validity to new methods that involve feature scaling and projection on a polyhedral strict subset of the feature subspace. Aside from greater generality and insight into the nature and effects of scaling, the potential benefits are implementation convenience (a feature matrix that need not have full rank), and possibly reduced overhead (no matrix inversion at each iteration). The comparison of methods within the class in terms of convergence rate, and the potential optimality of the projected Jacobi/LSPE-type method are open research questions.

ACKNOWLEDGMENT

This work was supported by NSF Grant ECCS-0801549.

REFERENCES


