Average Step Convergence of Undiscounted Linear-TD(0)

Tom Westerdale

October 14, 2023

Abstract

The linear-TD(0) method makes successive probabilistic adjustments in its estimates of state values. In a continuing task context, the step size is constant, and it is constant in this paper. For any step, we can calculate what the average would be of all the adjustments possible in that step, producing a notional average step. We show here that the sequence of such average steps converges. This convergence has already been proved for discounted linear-TD(0). This paper extends that proof to the undiscounted case, derives limit formulae for both cases, and shows why gradually removing the discount produces a formula different from the undiscounted limit formula. This paper plugs a gap in the foundations of adaptation analysis, allowing us to avoid a distorting discount if we wish.

1 Plan of the Paper

Linear-TD(0) methods are temporal difference methods that are used in adaptive systems to provide estimates of state values. In each time unit, the adaptive system takes an action, the environment changes its state, and payoff is received by the system. Payoff depends on state, and the value of the current state is a measure of how much payoff we expect in the future.¹

Linear-TD(0) uses a linear function

 $v_1\psi_{i1} + v_2\psi_{i2} + v_3\psi_{i3} + \dots + v_N\psi_{iN}$

to estimate the value of the current state i. The fixed numbers ψ_{ij} are values of features of state i. Linear-TD(0) iteratively modifies the coefficients v_j in an attempt to make the function a better estimator. In this paper, the step size is constant.

In defining state value, future payoff can be discounted by a small discount amount δ . This means that payoff *n* time steps in the future is multiplied by $(1-\delta)^n$. Linear-TD(0) can be discounted so that it estimates discounted value.

Discount Notation: Reinforcement Learning literature usually writes γ where I write $1-\delta$ and writes $1-\gamma$ where I write δ . Writing $1-\delta$ and δ as I do makes the tricky parts of the proofs here easier to understand. I will not use the symbol γ in this paper. Following everyday practice, I call δ the discount. (I will not use the phrase "discount rate".)

In this paper we discuss what I call average step convergence of linear-TD(0). Sutton and Barto outline a proof that discounted linear-TD(0) converges in the average step sense.[1, pp. 206-7] Here we extend that proof to the undscounted case. For comparison, we also repeat the shorter discounted case proof, giving it in our notation, and we exhibit the limit formulae for both cases.

We then look at what happens to the discounted case formula as $\delta \to 0$. At first sight we seem to obtain a sort of $\frac{0}{0}$ nonsense. In fact we do get a finite formula. It's different from the undiscounted case formula, but the difference makes sense.

We then look briefly at the value estimates, which I call false values because they can be very wrong. We write them in terms of what I call false payoffs.

Finally we finish up by briefly examining other versions of linear-TD(0).

¹Terminology in Adaptive Systems is inconsistent. We follow the usual Evolutionary Computation terminology. What we call payoff is what Reinforcement Learning researchers call reward, and what we call reward is what Reinforcement Learning researchers call reinforcement. What we call value is what many Evolutionary Adaptation researchers call fitness, yet the term fitness is often used to mean reproductive rate.

2 Definitions and Notation

2.1 State Value

In this paper, vectors are row vectors. Their transposes are column vectors. Each vector will be written as a bold lower case letter. Entries in the vector will be the corresponding italic letter. With one exception, each matrix will be a capital Latin letter. An entry in the matrix will be the same letter, but with a double subscript.

We have a strongly connected finite state Markov chain with N states. (N > 1)The $N \times N$ matrix P is the state transition probability matrix of the chain. So P is row stochastic. In this paper, P_{ij} is positive for every legal state transition $i \rightarrow j$.

It is the conditional probability that the next state is j given that the current state is i.

The vector \mathbf{e} is the vector that is simply a row of N ones. So $\mathbf{e} \mathbf{e}^{\top} = N$ and $P \mathbf{e}^{\top} = \mathbf{e}^{\top}$.

The vector $\tilde{\mathbf{p}}$ is the absolute state probability vector. So $\tilde{\mathbf{p}}P = \tilde{\mathbf{p}}$ and $\tilde{\mathbf{p}}\mathbf{e}^{\top} = 1$.

The absolute probability the chain is in state i is \tilde{p}_i . Every \tilde{p}_i is positive.

The matrix D is the diagonal matrix whose ii'th entry is \tilde{p}_i .

We define F = DP, so $F_{ij} = \tilde{p}_i P_{ij}$. I call F_{ij} the frequency of transition $i \to j$.

Associated with each state is a fixed real number called its payoff. When the chain enters a state, we receive the payoff associated with that state.

The vector \mathbf{m} is the vector of state payoffs. The vector \mathbf{m} doesn't change.

The scalar \bar{m} is the average payoff. That is, $\bar{m} = \tilde{\mathbf{p}} \mathbf{m}^{\top}$.

The vector **a** is the vector of *excess payoffs*. That is, $\mathbf{a} = \mathbf{m} - \bar{m} \mathbf{e}$.

A state's excess payoff is the amount by which its payoff exceeds the average payoff.

The average excess payoff is of course zero. $\mathbf{\tilde{p}} \mathbf{a}^{\top} = 0$.

The column vector \mathbf{c}^{\top} is the column vector of state values. Its definition is the Cesaro sum $\mathbf{c}^{\top} = \sum_{n=0}^{\infty} (P^n \mathbf{a}^{\top})$. The sum converges in the Cesaro sense.²

A discount is a non-negative real number δ less than 1. In defining state value, we can discount future excess payoff. The column vector of discounted state values is

$$\mathbf{c}_{\delta}^{\top} = \sum_{n=0}^{\infty} \left(\left(1 - \delta \right)^n P^n \mathbf{a}^{\top} \right) \quad . \tag{1}$$

If $\delta > 0$ then the Cesaro sum is the same as the ordinary sum.

If $\delta = 0$ then we say the values are undiscounted, and $\mathbf{c}_0^\top = \mathbf{c}^\top$.

2.2 Linear-TD(0) and its Average Step

There are \bar{N} basis functions $\psi_1, \psi_2, \psi_3, ..., \psi_{\bar{N}}$. Each ψ_k is a real valued function from the set of states. If i is a state then $\psi_k(i)$ is the value of its k'th feature. We write $\psi_k(i)$ as ψ_{ik} . The matrix Ψ is the $N \times \bar{N}$ matrix whose ik'th entry is ψ_{ik} . If the current state is i, then the vector of basis function values $\langle \psi_{i1}, \psi_{i2}, \psi_{i3}, ..., \psi_{i\bar{N}} \rangle$ is available to the system, and the current payoff m_i is available to the system.

To estimate the value of the current state, the system uses a vector $\mathbf{v} = \langle v_1, v_2, v_3, \dots, v_{\bar{N}} \rangle$ of real parameters that I call *cash balances*.³ The estimate \bar{v}_i of the value of state *i* is $\bar{v}_i = \sum_k v_k \psi_{ik}$. The column vector $\bar{\mathbf{v}}^{\top}$ of state value estimates is then given by $\bar{\mathbf{v}}^{\top} = \Psi \mathbf{v}^{\top}$.

I think of \overline{N} as being smaller than N, though it doesn't have to be. The vectors \mathbf{e} , $\mathbf{\tilde{p}}$, $\mathbf{\bar{v}}$, \mathbf{a} , and \mathbf{c} are what I call long vectors, because they have N entries. The vector \mathbf{v} is what I call a short vector, because it has only \overline{N} entries.

So \mathbf{e} is the long vector whose every entry is 1.

We define $\hat{\mathbf{e}}$ to be the short vector whose every entry is 1.

We also define \mathbf{e}_i to be the long vector whose *i*'th entry is 1 and whose other entries are 0.

And we define $\hat{\mathbf{e}}_{j}$ to be the short vector whose j'th entry is 1 and whose other entries are 0.

²For proof in this notation that it converges, see [3]. In this paper, I call c_i a state value, but elsewhere I've called it a *post-value* because the sum is only of payoffs that occur on or *after* the visit to the state.

³They are called cash balances in the bucket brigade, a special case of linear-TD(0). For convenience, I use the term cash balance in any linear-TD(0) method.

We will have big square $N \times N$ matrices and small square $\bar{N} \times \bar{N}$ matrices

The matrix I is the $N \times N$ identity matrix. The matrix I is the $\overline{N} \times \overline{N}$ identity matrix.

The only difference is the size of the symbol.

In each time step, linear-TD(0) adjusts the vector \mathbf{v}^{\top} of cash balances. the version of linear-TD(0) that we will concentrate on adjusts the cash balances in this way. If the state transition is $i \to w$, it simply adds

$$\varepsilon \left(a_i - \bar{v}_i + (1 - \delta) \,\bar{v}_w\right) \psi_{ik} \tag{2}$$

to each cash balance v_k .

The ε is a small constant positive real number that we call the step size.

The δ is the discount.

So given that the current cash balance vector is \mathbf{v} , the average change in v_k is

$$\varepsilon \sum_{iw} F_{iw}(a_i - \bar{v}_i + (1 - \delta) \, \bar{v}_w) \, \psi_{ik} \quad , \tag{3}$$

which we can write as $\varepsilon \, \hat{\mathbf{e}}_k \, \Psi^\top D \left(\mathbf{a}^\top - \bar{\mathbf{v}}^\top + (1 - \delta) \, P \, \bar{\mathbf{v}}^\top \right)$. We define

$$Y_{\delta} = D\left(I - (1 - \delta)P\right) \quad .$$

The average change in the column vector \mathbf{v}^{\top} of cash balances is $\varepsilon \Psi^{\top} D \left(\mathbf{a}^{\top} - \bar{\mathbf{v}}^{\top} + (1 - \delta) P \bar{\mathbf{v}}^{\top} \right)$, which we can write as $\varepsilon \Psi^{\top} D \mathbf{a}^{\top} - \varepsilon \Psi^{\top} Y_{\delta} \Psi \mathbf{v}^{\top}$.

Consider the sequence of column cash balance vectors $(\mathbf{v}^{(0)})^{\top}, (\mathbf{v}^{(1)})^{\top}, (\mathbf{v}^{(2)})^{\top}, (\mathbf{v}^{(3)})^{\top}, \dots$, where we derive $(\mathbf{v}^{(n+1)})^{\top}$ from $(\mathbf{v}^{(n)})^{\top}$ by taking the average step. That is,

$$(\mathbf{v}^{(n+1)})^{\top} = (I - \varepsilon \Psi^{\top} Y_{\delta} \Psi) (\mathbf{v}^{(n)})^{\top} + \varepsilon \Psi^{\top} D \mathbf{a}^{\top} \quad .$$
(4)

We ask whether this sequence converges, and if so, what it converges to. I call this convergence *average step* convergence.⁴ Average step convergence has already been shown for $\delta > 0$. (See [1, pp. 206-7].) This paper extends that result to the $\delta = 0$ case. It also relates the two cases. The relationship makes sense, but it's not trivial.

3 Basic Convergence Proof

3.1 Positive Definite Matrices

Terminology:

A vector is *tidy* if all its entries are the same. A vector is *messy* if it is not tidy.

The notion of positive definiteness is usually applied to symmetric real matrices. But like Rich Sutton, we apply that notion to all square real matrices, symmetric or not.

Positive Definite:

A square real matrix A is *positive definite* if for any nonzero real vector \mathbf{x} , the scalar $\mathbf{x}A\mathbf{x}^{\top}$ is positive.

Almost Positive Definite:

A square real matrix A is almost positive definite if for any real vector \mathbf{x} , the scalar $\mathbf{x}A\mathbf{x}^{\top}$ is positive if \mathbf{x} is messy, and is zero if \mathbf{x} is tidy.

 $^{^{4}}$ That's not what it's called in the literature, but I find the terminology in the literature confusing.

Theorem 1

If A is a square real matrix that is positive definite, then all its eigenvalues are in the positive half plane.

Proof:

Suppose A is a real positive definite matrix and suppose z is a left eigenvector of A with eigenvalue λ .

Then there are two real vectors \mathbf{x} and \mathbf{y} , not both zero, such that $\mathbf{z} = \mathbf{x} + i\mathbf{y}$. $(i = \sqrt{-1})$ Define $\mathbf{z}^* = (\mathbf{x} - i\mathbf{y})^\top$. Then $\mathbf{z} \mathbf{z}^*$ is a positive real number. We have these real parts. $(\operatorname{Re}(\lambda))\mathbf{z}\,\mathbf{z}^* = \operatorname{Re}(\lambda\,\mathbf{z}\,\mathbf{z}^*) = \operatorname{Re}(\mathbf{z}A\,\mathbf{z}^*) = \mathbf{x}A\,\mathbf{x}^\top + \mathbf{y}A\,\mathbf{y}^\top.$ This is a positive real since both terms are non-negative and at least one is positive. So $\operatorname{Re}(\lambda) > 0$.

So positive definite matrices are nonsingular, since zero is not an eigenvalue.

3.2Undiscounted Y is Almost Positive Definite.

Notation: In this subsection all vectors are long vectors.

Consider a cycle of legal transitions that cycles through all the states and returns to the starting state, and in which each state is visited exactly once. Since the Markov chain is strongly connected, we know there is such a cycle. We select such a cycle and call it the *designated cycle* of transitions.

We call call a real square matrix A nice just if these conditions all hold.

- (1)Every diagonal entry of A is positive.
- Every off-diagonal entry of A is either negative or zero. (2)
- For every transition $i \to j$ in the designated cycle, the entry A_{ij} is negative. (3)
- $A\mathbf{e}^{\top} = 0$. (4) $\mathbf{e}A = 0$ and

We define Y = D(I - P). In this subsection only, we define $S = Y + Y^\top .$

Statements (1), (2), and (3) hold when A is I - P. So matrix Y is nice, and matrix S is nice.

We select an eigenvector **y** of S with eigenvalue λ . Since S is symmetric, eigenvalue λ is real. Let y_k be an entry in y whose modulus is not exceeded by the modulus of any other entry in y. Define $\mathbf{z} = y_k^{-1} \mathbf{y}$.

Three properties of the vector **z**

- Every entry in \mathbf{z} has modulus less than or equal to 1. (1)
- (2)At least one entry in \mathbf{z} is the number 1.
- \mathbf{z} is an eigenvector of S with eigenvalue λ . (3)

Lemma 1

If for distinct states i and k we have

 $z_i = 1$ and $S_{ik} \neq 0$, then:

(1) $\lambda \geq 0$,

 $\lambda = 0 \implies z_k = 1$. (2)

Proof:

Suppose for distinct states i and k we have $z_i = 1$ and $S_{ik} \neq 0$. We will use the three \mathbf{z} properties. And since S is nice, we will use the four nice properties for S. We have $\sum_{j} S_{ij} = \mathbf{e}_i \hat{S} \mathbf{e}^{\top} = 0$, so

$$S_{ii} = \sum_{j \neq i} (-S_{ij}) \quad . \tag{5}$$

The sum is over all j that are not equal to i. We also have $\sum_j S_{ij} z_j = \mathbf{e}_i S \mathbf{z}^\top = \mathbf{e}_i (\lambda \mathbf{z}^\top) = \lambda z_i = \lambda$. Therefore,

$$S_{ii} - \lambda = \sum_{j \neq i} (-S_{ij}) z_j \tag{6}$$

We have $\left|\sum_{j\neq i}(-S_{ij})z_{j}\right| \leq \sum_{j\neq i}(-S_{ij})|z_{j}| \leq \sum_{j\neq i}(-S_{ij})$. By (5) and (6) this is $|S_{ii} - \lambda| \leq S_{ii}$. Since S_{ii} is positive, we have $\lambda \geq 0$. Now suppose $\lambda = 0$. Write x_{j} for the real part of z_{j} . Then the right sides of (5) and (6) are

equal, so $\sum_{j \neq i} (-S_{ij})(1-z_j) = 0$, and $\sum_{j \neq i} (-S_{ij})(1-x_j) = 0$. Since $x_j \leq |z_j| \leq 1$, every

term in the last sum is non-negative. Hence every term is zero. In particular, $(-S_{ik})(1-x_k) = 0$, so $x_k = 1$. This and $|z_k| \le 1$ give us $z_k = 1$.

We can always select a z_i that equals 1 and select the corresponding transition $i \to k$ on the designated cycle. Then $S_{ik} \neq 0$, and lemma 1 tells us that $\lambda \geq 0$.

Suppose $\lambda = 0$. Select a transition $i \to k$ on the designated cycle such that $z_i = 1$. Then Lemma 1 tells us that $z_k = 1$. We now look at the next transition in the cycle and proceed in this way step by step around the cycle. So every entry z_j equals 1, and $\mathbf{z} = \mathbf{e}$. So we see that $\lambda = 0 \implies \mathbf{z} = \mathbf{e}$.

Conversely, suppose $\mathbf{z} = \mathbf{e}$. Then $\mathbf{z}S = \mathbf{e}S = 0$, so $\lambda = 0$. So we have these results.

 $\lambda = 0$ if and only if $\mathbf{z} = \mathbf{e}$.

 $\lambda = 0$ if and only if **y** is tidy.

The tidy eigenvectors of S have zero eigenvalues, and

the messy eigenvectors of S have positive eigenvalues.

Since S is symmetric, there is a set of orthogonal real eigenvectors $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \dots, \mathbf{z}_N$ that span the space. I'll call that set of eigenvectors our basis, and I'll write λ_k for the eigenvalue of \mathbf{z}_k . Now \mathbf{e} is an eigenvector with eigenvalue 0, so since S is symmetric, if $\lambda_k > 0$ then \mathbf{z}_k is orthogonal to \mathbf{e} . So the basis vectors can't all be messy since then they would all be orthogonal to \mathbf{e} and wouldn't span the space. So there is a tidy basis vector, and obviously it must be the only one. Let's re-order the basis vectors so \mathbf{z}_1 is the tidy vector.

Now take an arbitrary real vector and write it $\mathbf{x} = \sum_k b_k \mathbf{z}_k$. By taking real parts we make every scalar b_k real. Then we have $\mathbf{x} S \mathbf{x}^{\top} = \sum_k b_k^2 \lambda_k \mathbf{z}_k \mathbf{z}_k^{\top}$,

since the other terms are zero.

The eigenvalues λ_k are non-negative so every term in the sum is non-negative, and so is $\mathbf{x} S \mathbf{x}^{\top}$.

Suppose $\mathbf{x} S \mathbf{x}^{\top}$ is zero. Then every term in the sum must be zero. Now λ_1 is zero, but the other eigenvalues λ_k are positive, so every b_k is zero except possibly b_1 . So we see that $\mathbf{x} = b_i \mathbf{z}_1$, and so \mathbf{x} is tidy.

Conversely, if \mathbf{x} is tidy then $\mathbf{x} S \mathbf{x}^{\top}$ is zero, since $\mathbf{e}S = 0$. So we see that S is almost positive definite.

Since $\mathbf{x} S \mathbf{x}^{\top} = 2 \mathbf{x} Y \mathbf{x}^{\top}$, we conclude:

Y is almost positive definite.

3.3 Discounted Y_{δ} is Positive Definite.

Note that $Y_{\delta} = D(I - (1 - \delta)P) = Y + \delta F$ and $Y_0 = Y$. In this subsection only, we define $S = Y_{\delta} + Y_{\delta}^{\top}$. Both Y_{δ} and S satisfy statements (1), (2), and (3) in the definition of a nice matrix. $\mathbf{e} S = 2\delta \mathbf{\tilde{p}} \qquad S \mathbf{e}^{\top} = 2\delta \mathbf{\tilde{p}}^{\top}$

Now suppose $\delta > 0$.

We now look at the eigenvalues of S. Since S is symmetric, the eigenvalues are real. Let's look at the i'th row of S.

$$\begin{split} \sum_{j} S_{ij} &= \mathbf{e}_{i} S \mathbf{e}^{\top} = \mathbf{e}_{i} (2\delta \, \tilde{\mathbf{p}}^{\top}) = 2\delta \, \tilde{p}_{i} > 0 \\ S_{ii} &> \sum_{j \neq i} (-S_{ij}) \qquad \text{(The sum is over all } j \text{ that are not equal to } i.) \\ S_{ii} &> \sum_{j \neq i} |S_{ij}| \\ \text{The Conversion dials for a state of the sum is over all } j \text{ that are not equal to } i.) \end{split}$$

The Geršgorin disk for row i is the closed disk in the complex plane whose center is S_{ii} and whose radius is $\sum_{j \neq i} |S_{ij}|$. We see by the last inequality that every Geršgorin disk is entirely within the positive half plane. The Geršgorin set is the union of the Geršgorin disks, and we see that it is entirely within the positive half plane. By Geršgorin's theorem, every eigenvalue is in the Geršgorin set. So every eigenvalue of S is in the positive half plane. Every eigenvalue is a positive real.

Since S is symmetric, there is a set of real orthogonal eigenvectors $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \dots, \mathbf{z}_{\bar{N}}$ that span the space. I'll call that set of eigenvectors our basis. I'll write λ_k for the eigenvalue of \mathbf{z}_k .

Now let \mathbf{x} be an arbitrary nonzero real vector. We write it in terms of the basis.

 $\mathbf{x} = \sum_k b_k \mathbf{z}_k$

The x and each \mathbf{z}_k are real, so we can take real parts of everything and end up with the above equation, but with every b_k real. So now every b_k is real. Since x is nonzero, at least one b_k is nonzero. We have $\mathbf{x} S \mathbf{x}^{\top} = \sum_{k} b_k^2 \lambda_k \mathbf{z}_k \mathbf{z}_k^{\top}$, since the other terms are zero. For any k we have the following.

 b_k^2 is non-negative, since b_k is real.

 λ_k is a positive real.

 $\mathbf{z}_k \mathbf{z}_k^{\top}$ is a positive real, since \mathbf{z}_k is real and nonzero. Furthermore, at least one b_k is nonzero, so that b_k^2 is positive.

So we see that

 $\mathbf{x} S \mathbf{x}^\top > 0 \; .$

The \mathbf{x} was an arbitrary nonzero vector, so S is positive definite.

Since $\mathbf{x} S \mathbf{x}^{\top} = 2 \mathbf{x} Y_{\delta} \mathbf{x}^{\top}$ we conclude:

If $\delta > 0$ then Y_{δ} is positive definite.

3.4 Happy and Sad Vectors

If we have a complex vector \mathbf{z} , we can of course write it as $\mathbf{w} + i\mathbf{r}$, where \mathbf{w} and \mathbf{r} are real vectors (and i is $\sqrt{-1}$). We call w the real part of z, and we call r the imaginary part of z.

We define these sets of short vectors.

 $\mathcal{K} = \{ \mathbf{y} | \, \mathbf{y} \, \Psi^\top = 0 \}$ $\mathcal{H} = \{\mathbf{y} | \mathbf{y} \Psi^{\top} \text{ is tidy}\}$

I call \mathcal{K} the kernel and call its members kernel vectors. I call the members of \mathcal{H} happy vectors.

It's easy to see that if \mathbf{y} is a happy vector then its real and imaginary parts are both happy. And if \mathbf{y} is a kernel vector then its real and imaginary parts are both kernel vectors.

We see that \mathcal{H} and \mathcal{K} are both subspaces and that $\mathcal{K} \subseteq \mathcal{H}$.

Let's look at Ψ^{\top} as a linear transformation from \mathcal{H} . So \mathcal{H} is its domain, \mathcal{K} is its kernel, and its range is a subspace of tidy vectors. If $\mathcal{H} \neq \mathcal{K}$ then there is a nonzero tidy vector in the range, every tidy vector is in the range, and the dimension of the range is 1. So the dimension of \mathcal{H} is one more than the dimension of \mathcal{K} .

 $\mathcal{H} = \mathcal{K}$ or the dimension of \mathcal{H} is one more than the dimension of \mathcal{K} . Either

 $\mathcal{H} \neq \mathcal{K}$ if and only if **e** is in the range of Ψ^{\top} . Note that

To say that vectors \mathbf{z} and \mathbf{w} are orthogonal means of course that $\mathbf{z} \mathbf{w}^* = 0$, where * means complex conjugate transpose. If at least one of the two vectors is real, then they are orthogonal if and only if $\mathbf{z} \mathbf{w}^{\top} = 0$. Usually when I refer to two vectors being orthogonal, one of the vectors will be real. I shall sometimes use the special term orthogonal^{*}. Vectors \mathbf{z} and \mathbf{w} are orthogonal^{*} if they are orthogonal and at least one of them is real. The * is simply a reminder that because one of them is real, orthogonality in this case is equivalent to $\mathbf{z} \mathbf{w}^{\top} = 0$.

The norm $\|\mathbf{w}\|$ of vector \mathbf{w} is of course the square root of $\mathbf{w}\mathbf{w}^*$. If \mathbf{w} is real then its norm is the square root of $\mathbf{w} \mathbf{w}^{\top}$. To say that real \mathbf{w} has norm 1 is to say that $\mathbf{w} \mathbf{w}^{\top} = 1$. The norm of a vector I sometimes call its length.

Lemma 2

The subspace \mathcal{H} has an orthonormal basis in which every basis vector is real.

Proof:

By induction.

Here is the induction step.

Suppose there is an orthonormal set of real vectors $\mathcal{E} = \{\eta_1, \eta_2, \eta_3, ..., \eta_\ell\}$ that is a subset of \mathcal{H} but does not span \mathcal{H} . Using the Gram-Schmidt process, we obtain a nonzero member \mathbf{z} of \mathcal{H} that is orthogonal* to every $\ \boldsymbol{\eta}_i$ in $\ \mathcal{E}$.

The real part and the imaginary part of \mathbf{z} are both in \mathcal{H} . The real part is either zero or it's orthogonal^{*} to every member of \mathcal{E} . The same is true for the imaginary part. They can't both be zero, so select a nonzero part. It's a real vector in $\mathcal H$ that is orthogonal^{*} to every member of $\mathcal E$. Now adjust its length to 1.

By virtually the same proof, \mathcal{K} also has an orthonormal basis in which every basis vector is real.

Now suppose $\mathcal{H} \neq \mathcal{K}$. Look at the induction step in the proof of lemma 2, and let $\{\eta_1, \eta_2, \eta_3, ..., \eta_\ell\}$ be a real orthonormal basis of \mathcal{K} . The induction step gives us a new basis vector $\eta_{\ell+1}$, and adding that vector to the basis gives us a basis of the whole of \mathcal{H} . We see that $\eta_{\ell+1} \Psi^{\top}$ is a nonzero tidy real vector, so $\eta_{\ell+1} \Psi^{\top} = \beta \mathbf{e}$, where β is a nonzero real scalar. If β is negative, let's change $\eta_{\ell+1}$, multiplying it by -1. It's still just as good a basis vector, and now β is positive. We shall use an \mathcal{H} basis constructed in this way.

In what follows, I shall use the letter η to mean the vector in our basis of \mathcal{H} that is not in \mathcal{K} . If $\mathcal{H} \neq \mathcal{K}$ then

 $\boldsymbol{\eta} \boldsymbol{\Psi}^{\mathsf{T}} = \boldsymbol{\beta} \mathbf{e} \; .$

 $\beta > 0$

If $\mathcal{H} = \mathcal{K}$ then there is no vector $\boldsymbol{\eta}$ and no scalar β .

Let $\mathcal{E} = \{\eta_1, \eta_2, \eta_3, ..., \eta_\ell\}$ be our orthonormal basis of \mathcal{H} . We define the real symmetric matrix H.

$$H = \sum_{j=1}^{\ell} \boldsymbol{\eta}_j^{\mathsf{T}} \boldsymbol{\eta}_j \quad . \tag{7}$$

If $\mathcal{E} = \emptyset$ (\mathcal{H} is a singleton) then the matrix H has every entry zero.

Note that $\eta_j H = \eta_j$ for any basis vector η_j . Therefore, since H is a linear transformation we have this. If $\mathbf{y} \in \mathcal{H}$, then $\mathbf{y} H = \mathbf{y}$.

We also define \pm

 $K = \sum_{j} \boldsymbol{\eta}_{j}^{\dagger} \boldsymbol{\eta}_{j},$

where in the sum we use just the vectors in the \mathcal{K} basis. Then just as for \mathcal{H} , we have

if $\mathbf{y} \in \mathcal{K}$ then $\mathbf{y} \mathbf{K} = \mathbf{y}$. We also have these.

If $\mathcal{H} = \mathcal{K}$ then H = K.

If $\mathcal{H} \neq \mathcal{K}$ then $H = K + \eta^{\top} \eta$,

where η is of course the basis vector that is not in the basis of \mathcal{K} .

Both H and K are symmetric, so we have this.

If **y** is a happy vector then $\mathbf{y}H = \mathbf{y}$ and $H\mathbf{y}^{\top} = \mathbf{y}^{\top}$. If **y** is a kernel vector then $\mathbf{y}K = \mathbf{y}$ and $K\mathbf{y}^{\top} = \mathbf{y}^{\top}$.

We call a short vector sad if it is orthogonal to every happy vector. Equivalently, a vector is sad if it is orthogonal^{*} to every basis vector η_k in \mathcal{E} . Let \mathcal{S} be the subspace of sad vectors.

We see that if \mathbf{x} is a sad vector then $\mathbf{x} H = 0$ and $H \mathbf{x}^{\top} = 0$.

Suppose **x** is a short vector. We have $\mathbf{x} H \Psi^{\top} = \mathbf{x} \left(\sum_{k} \boldsymbol{\eta}_{k}^{\top} \boldsymbol{\eta}_{k} \right) \Psi^{\top} = \sum_{k} (\mathbf{x} \boldsymbol{\eta}_{k}^{\top}) (\boldsymbol{\eta}_{k} \Psi^{\top})$. We see that each $(\mathbf{x} \boldsymbol{\eta}_{k}^{\top})$ is a complex scalar and each $(\boldsymbol{\eta}_{k} \Psi)$ is a tidy vector, so $\mathbf{x} H \Psi^{\top}$ is a tidy vector and $\mathbf{x} H$ is a happy vector.

We also have $(I - H) \mathbf{y}^{\top} = 0$ for any happy vector \mathbf{y} , so $\mathbf{x} (I - H) \mathbf{y}^{\top} = 0$ for any happy vector \mathbf{y} . Therefore, $\mathbf{x} (I - H)$ is a sad vector.

We call $\mathbf{x}H$ the happy part of \mathbf{x} , and we call $\mathbf{x}(I-H)$ the sad part of \mathbf{x} . Suppose we write \mathbf{x} as the sum $\mathbf{z} + \mathbf{y}$ of a sad vector \mathbf{z} and a happy vector \mathbf{y} . Then $\mathbf{x}H = \mathbf{y}$ and $\mathbf{x}(I-H) = \mathbf{z}$. The happy part is \mathbf{y} and the sad part is \mathbf{z} . A short vector can be written as the sum of a sad vector and a happy vector in one and only one way. We see that if a short vector is real then its sad part and happy part are both real.

We define the subspace of antikernel vectors. A vector is an antikernel vector just if it is orthogonal to every kernel vector. Every short vector can be written as the sum of an antikernel vector and a kernel vector in one and only one way. If \mathbf{x} is a short vector then $\mathbf{x}K$ is its kernel part and $\mathbf{x}(I-K)$ is its antikernel part. Note that if \mathbf{z} is an antikernel vector then $\mathbf{z}K = 0$ and $K\mathbf{z}^{\top} = 0$.

3.5 B_0 and A_{δ} are Positive Definite

We define

 $A_{\delta} = \Psi^{\top} Y_{\delta} \Psi + K$

 $B_{\delta} = \Psi^{\top} Y_{\delta} \Psi + H$

In particular, we have

 $B_0 = \Psi^\top Y \Psi + H .$

If $\mathbf{y} \in \mathcal{K}$ then $A_{\delta} \mathbf{y}^{\top} = \mathbf{y}^{\top}$. If $\mathbf{y} \in \mathcal{H}$ then $B_0 \mathbf{y}^{\top} = \mathbf{y}^{\top}$.

In this subsection, \mathbf{x} will be a nonzero real short vector.

We have

 $\mathbf{x} \, H \, \mathbf{x}^\top = \sum_j (\mathbf{x} \, \boldsymbol{\eta}_j^\top) (\boldsymbol{\eta}_j \mathbf{x}^\top) \geq 0$. Furthermore, if \mathbf{x} is happy, then $\mathbf{x} H \mathbf{x}^{\top} = \mathbf{x} \mathbf{x}^{\top} > 0 .$

Now Y is almost positive definite, so $\mathbf{x} \Psi^{\top} Y \Psi \mathbf{x}^{\top}$ is non-negative.

Furthermore, if **x** is not happy then $\mathbf{x} \Psi^{\top}$ is messy and $\mathbf{x} \Psi^{\top} Y \Psi \mathbf{x}^{\top}$ is positive.

The scalar $\mathbf{x} B_0 \mathbf{x}^{\top}$ can be written as the following sum.

 $\mathbf{x} \Psi^{\top} Y \Psi \mathbf{x}^{\top} + \mathbf{x} H \mathbf{x}^{\top}$

Both terms are non-negative. If \mathbf{x} is happy then the second term is positive, and if \mathbf{x} is not happy then the first term is positive. So in any case the sum is positive. We conclude:

 B_0 is positive definite.

Also, we have $\mathbf{x} K \mathbf{x}^{\top} = \sum_{k} (\mathbf{x} \boldsymbol{\eta}_{k}^{\top}) (\boldsymbol{\eta}_{k} \mathbf{x}^{\top}) \geq 0$. If \mathbf{x} is in \mathcal{K} then $\mathbf{x} K \mathbf{x}^{\top} = \mathbf{x} \mathbf{x}^{\top} > 0$. Now suppose $\delta > 0$. Then Y_{δ} is positive definite. So $\mathbf{x} \Psi^{\top} Y_{\delta} \Psi \mathbf{x}^{\top}$ is non-negative, and furthermore, if $\mathbf{x} \notin \mathcal{K}$ then $\mathbf{x} \Psi^{\top} Y_{\delta} \Psi \mathbf{x}^{\top}$ is positive. The scalar $\mathbf{x} A_{\delta} \mathbf{x}^{\top}$ can be written as the following sum.

$$\mathbf{x} \Psi^{\top} Y_{\delta} \Psi \mathbf{x}^{\top} + \mathbf{x} K \mathbf{x}^{\top}$$

Both terms are non-negative. If $\mathbf{x} \in \mathcal{K}$ then the second term is positive, and if $\mathbf{x} \notin \mathcal{K}$ then the first term is positive. So in any case the sum is positive. We conclude:

If $\delta > 0$ then A_{δ} is positive definite.

Still assume $\delta > 0$.

If $\mathcal{H} \neq \mathcal{K}$ then $B_{\delta} = A_{\delta} + \boldsymbol{\eta}^{\top} \boldsymbol{\eta}$, where η is the happy basis vector that isn't in \mathcal{K} . $\mathbf{x} B_{\delta} \mathbf{x}^{\dagger} = \mathbf{x} A_{\delta} \mathbf{x}^{\dagger} + (\mathbf{x} \boldsymbol{\eta}^{\dagger}) (\boldsymbol{\eta} \mathbf{x}^{\dagger})$

The second term on the right is non-negative, and A_{δ} is positive definite,

so B_{δ} is positive definite when $\delta > 0$. When $\delta = 0$ we have $B_{\delta} = B_0$, and we said that's positive definite. So whether δ is zero or not,

 B_{δ} is positive definite.

Remember that positive definite matrices are nonsingular. We shall see that in some cases A_0 is singular, and this makes our story complicated.

The Undiscounted Case $\mathbf{3.6}$

So by theorem 1, all the eigenvalues of B_0 are in the positive half plane. So they are all inside some circle tangent to the imaginary axis at the origin. We are going to move that circle. We define

 $Q = I - \varepsilon B_0 ,$

where ε is a small positive real.

We see that if ε is small enough, the the eigenvalues of Q will all be inside the unit circle.

We insist that ε be that small or smaller. The ε will be our step size.

So all the eigenvalues of Q are inside the unit circle.

The spectral radius of Q is less than 1.

Note that $B_0 = \frac{1}{\varepsilon} (I - Q)$, so since B_0 is nonsingular we have $B_0^{-1} = \varepsilon (I - Q)^{-1}$.

Note that if \mathbf{y} is happy then $Q \mathbf{y}^{\top} = (1 - \varepsilon) \mathbf{y}^{\top}$.

Let η_k be a basis vector of \mathcal{H} . Then $\eta_k \Psi^{\top}$ is a tidy vector, so
$$\begin{split} \boldsymbol{\eta}_k \boldsymbol{\Psi}^\top \boldsymbol{Y} &= 0 \quad \text{and} \quad \boldsymbol{\eta}_k \boldsymbol{\Psi}^\top \boldsymbol{D} \, \mathbf{a}^\top = 0 \ . \\ \text{Therefore,} \quad \boldsymbol{H} \, \boldsymbol{\Psi}^\top \boldsymbol{Y} \,= \, 0 \quad \text{and} \quad \boldsymbol{H} \, \boldsymbol{\Psi}^\top \boldsymbol{D} \, \mathbf{a}^\top = \, 0 \ . \end{split}$$

We look at the sequence of successive cash balance column vectors.

 $(\mathbf{v}^{(0)})^{\top}, \ (\mathbf{v}^{(1)})^{\top}, \ (\mathbf{v}^{(2)})^{\top}, \ (\mathbf{v}^{(3)})^{\top}, \ \dots$ The average step equation is this. $(\mathbf{v}^{(n+1)})^{\top} = (I - \varepsilon \Psi^{\top} Y \Psi) (\mathbf{v}^{(n)})^{\top} + \varepsilon \Psi^{\top} D \mathbf{a}^{\top}$ Multiplying by H on the left gives us $H(\mathbf{v}^{(n+1)})^{\top} = H(\mathbf{v}^{(n)})^{\top}$. So the happy part of the column vectors is unchanged along the sequence, so for each n we can write $(\mathbf{v}^{(n)})^{\top} = (\mathbf{x}^{(n)})^{\top} + \mathbf{y}^{\top}$, where $(\mathbf{x}^{(n)})^{\top}$ is the sad part and \mathbf{y}^{\top} is the common happy part. We also have $I - \varepsilon \Psi^\top Y \Psi = Q + \varepsilon H$. Making these substitutions in the average step equation gives us this. $\begin{aligned} & (\mathbf{x}^{(n+1)})^{\top} + \mathbf{y}^{\top} = (Q + \varepsilon H) \left((\mathbf{x}^{(n)})^{\top} + \mathbf{y}^{\top} \right) + \varepsilon \Psi^{\top} D \mathbf{a}^{\top} \\ & \text{We use} \quad Q \mathbf{y}^{\top} = (1 - \varepsilon) \mathbf{y}^{\top} \quad \text{and} \quad H \mathbf{y}^{\top} = \mathbf{y}^{\top} \quad \text{and} \quad H (\mathbf{x}^{(n)})^{\top} = 0 . \end{aligned}$ We obtain $(\mathbf{x}^{(n+1)})^{\top} = Q (\mathbf{x}^{(n)})^{\top} + \varepsilon \Psi^{\top} D \mathbf{a}^{\top} \end{aligned}$ By induction on n we have $(\mathbf{x}^{(n)})^{\top} = \varepsilon \left(\sum_{k=0}^{n-1} Q^k \right) \Psi^{\top} D \, \mathbf{a}^{\top} + Q^n (\mathbf{x}^{(0)})^{\top} \ .$ We saw that $(I-Q)^{-1}$ exists, so the equation becomes $(\mathbf{x}^{(n)})^{\top} = \varepsilon \left(I - Q^n\right) \left(I - Q\right)^{-1} \Psi^{\top} D \, \mathbf{a}^{\top} + Q^n (\mathbf{x}^{(0)})^{\top} \, .$ Since the spectral radius of Q is less than 1, we have $\lim_{n\to\infty} (\mathbf{x}^{(n)})^{\top} = \varepsilon (I-Q)^{-1} \Psi^{\top} D \mathbf{a}^{\top} = B_0^{-1} \Psi^{\top} D \mathbf{a}^{\top}.$ $\lim_{n \to \infty} Q^n = 0$ and We define

$$\mathbf{u}^{\top} = B_0^{-1} \Psi^{\top} D \mathbf{a}^{\top}$$

If we begin with $(\mathbf{v}^{(0)})^{\top} = \mathbf{x}^{\top} + \mathbf{y}^{\top}$, where \mathbf{x}^{\top} is the sad part and \mathbf{y}^{\top} is the happy part, then the limit is $\mathbf{u}_{\top}^{\top} + \mathbf{y}^{\top}$.

We note that \mathbf{u}^{\top} is sad. We can easily show this directly. Let \mathbf{y} be any happy vector. Then there is a complex scalar α such that $\mathbf{y} \Psi^{\top} = \alpha \mathbf{e}$. We have $\mathbf{y} \Psi^{\top} Y = \alpha \mathbf{e} Y = 0$, so we have $\mathbf{y} B_0 = \mathbf{y}$ and $\mathbf{y} B_0^{-1} = \mathbf{y}$. Therefore, $\mathbf{y} \mathbf{u}^{\top} = \mathbf{y} B_0^{-1} \Psi^{\top} D \mathbf{a}^{\top} = \mathbf{y} \Psi^{\top} D \mathbf{a}^{\top} = \alpha \mathbf{e} D \mathbf{a}^{\top} = 0$.

3.7The Discounted Case

Now assume $\delta > 0$.

We now proceed just as in the undiscounted case. We define $Q_{\delta} = I - \varepsilon A_{\delta}$, and we insist that ε be small enough that the spectral radius of Q_{δ} is less than 1.

 $A_{\delta}^{-1} = \varepsilon \left(I - Q_{\delta} \right)^{-1}$ If $\mathbf{y} \in \mathcal{K}$ then $Q_{\delta} \mathbf{y}^{\top} = (1 - \varepsilon) \mathbf{y}^{\top}$.

We look at the sequence of successive cash balance column vectors. $(\mathbf{v}^{(0)})^{\top}, (\mathbf{v}^{(1)})^{\top}, (\mathbf{v}^{(2)})^{\top}, \dots$ The average step equation is this.

 $(\mathbf{v}^{(n+1)})^{\top} = (I - \varepsilon \Psi^{\top} Y_{\delta} \Psi) (\mathbf{v}^{(n)})^{\top} + \varepsilon \Psi^{\top} D \mathbf{a}^{\top}$

Multiplying by K and using $K\Psi^{\top} = 0$ gives us $K(\mathbf{v}^{(n+1)})^{\top} = K(\mathbf{v}^{(n)})^{\top}$. The kernel part is constant along the sequence. We write $(\mathbf{v}^{(n)})^{\top} = (\mathbf{x}^{(n)})^{\top} + \mathbf{y}^{\top}$, where $(\mathbf{x}^{(n)})^{\top}$ is the antikernel part and \mathbf{y}^{\top} is the common kernel part. Making this substitution and $I - \varepsilon \Psi^{\top} Y_{\delta} \Psi = Q_{\delta} + \varepsilon K$ gives us $(\mathbf{x}^{(n+1)})^{\top} = Q_{\delta} (\mathbf{x}^{(n)})^{\top} + \varepsilon \Psi^{\top} D \mathbf{a}^{\top}.$

Just as in the undiscounted case, we inductively obtain a formula for $(\mathbf{x}^{(n)})^{\top}$ and then let $n \to \infty$. We obtain $\lim_{n\to\infty} (\mathbf{x}^{(n)})^{\top} = \mathbf{u}_{\delta}^{\top}$, where $\mathbf{u}_{\delta}^{\top} = \varepsilon (I - Q_{\delta})^{-1} \Psi^{\top} D \mathbf{a}^{\top}$. u

$$\mathbf{u}_{\delta}^{ op} = A_{\delta}^{-1} \Psi^{ op} D \mathbf{a}^{ op}$$
 .

Of course this limit holds only if $\delta > 0$.

Decreasing δ 4

We now ask what happens to $\mathbf{u}_{\delta}^{\top}$ if $\delta \to 0$.

The matrix B_{δ} is nonsingular for $\delta \geq 0$, and the inverse of a nonsingular matrix is a continuous function of the matrix.⁵ So B_{δ}^{-1} is a continuous function of δ , and $B_{\delta}^{-1} \to B_{0}^{-1}$ as $\delta \to 0$. If $\mathcal{H} = \mathcal{K}$ then $A_{\delta} = B_{\delta}$ and $\mathbf{u}_{\delta}^{\top} = B_{\delta}^{-1} \Psi^{\top} D \mathbf{a}^{\top}$, so obviously $\mathbf{u}_{\delta}^{\top} \to \mathbf{u}^{\top}$.

So in the rest of this section we will assume $\mathcal{H} \neq \mathcal{K}$. Now what happens to $\mathbf{u}_{\delta}^{\top}$?

⁵If matrix A is nonsingular then $A^{-1} = |A|^{-1} \operatorname{adj}(A)$. By $\operatorname{adj}(A)$ I mean the adjugate of A.

In that case we have $H = K + \boldsymbol{\eta}^{\top} \boldsymbol{\eta} ,$ $\boldsymbol{\eta} \Psi^{\top} = \beta \mathbf{e} ,$ $\beta > 0 .$ $B_{\delta} = A_{\delta} + \boldsymbol{\eta}^{\top} \boldsymbol{\eta}$

4.1 Tiniest Eigenvalue of A_{δ}

We define two subspaces:

 $\begin{array}{l} \mathcal{Y} \text{ is the vectors that are scalar multiples of } \boldsymbol{\eta} \text{ .} \\ \mathcal{B} \text{ is the vectors that are orthogonal* to } \boldsymbol{\eta} \text{ .} \\ \text{Since } \boldsymbol{\eta} \text{ is happy, we have these simple facts.} \\ \boldsymbol{\eta} K = 0 \qquad K \boldsymbol{\eta}^{\top} = 0 \qquad \boldsymbol{\eta} H = \boldsymbol{\eta} \qquad H \boldsymbol{\eta}^{\top} = \boldsymbol{\eta}^{\top} \qquad \boldsymbol{\eta} \Psi^{\top} Y = 0 \qquad Y \Psi \boldsymbol{\eta}^{\top} = 0 \\ \text{Therefore we have these.} \\ \boldsymbol{\eta} A_0 = 0 \qquad A_0 \boldsymbol{\eta}^{\top} = 0 \qquad \boldsymbol{\eta} B_0 = \boldsymbol{\eta} \qquad B_0 \boldsymbol{\eta}^{\top} = \boldsymbol{\eta}^{\top} \\ \text{If } \mathbf{x} \in \mathcal{B} \quad \text{then } (\mathbf{x} B_0) \boldsymbol{\eta}^{\top} = \mathbf{x} (B_0 \boldsymbol{\eta}^{\top}) = \mathbf{x} \boldsymbol{\eta}^{\top} = 0 \text{ , so } \qquad \mathbf{x} B_0 \in \mathcal{B} \text{ .} \\ \text{So we see that the linear transformation } B_0 \text{ maps } \mathcal{B} \text{ into itself,} \\ \text{and also maps } \mathcal{Y} \text{ into itself. In fact, it is the identity transformation on } \mathcal{Y} \text{ .} \end{array}$

Furthermore, if $\mathbf{x} \in \mathcal{B}$ we have $\mathbf{x}B_0 = \mathbf{x}A_0 + \mathbf{x}\boldsymbol{\eta}^\top\boldsymbol{\eta} = \mathbf{x}A_0$.

Transformation A_0 agrees with B_0 on the subspace \mathcal{B} .

Transformation A_0 maps the whole \mathcal{Y} subspace to the zero vector.

Consider a basis in which η is the first basis element and in which all the other basis elements are members of \mathcal{B} . Let's write B_0 and A_0 using that basis. The matrix B_0 is block diagonal with just two blocks, a 1×1 block and an $(\bar{N}-1) \times (\bar{N}-1)$ block. The 1×1 block is just the number 1. The matrix A_0 is similarly block diagonal with the same $(\bar{N}-1) \times (\bar{N}-1)$ block, but here the 1×1 block is just the number 0.

The eigenvalues of B_0 are the eigenvalues of the big block plus the number 1. The eigenvalues of A_0 are the same big block eigenvalues plus the number 0. Matrix B_0 is nonsingular, so none of its eigenvalues are zero. So none of the big block eigenvalues are zero. Therefore, 0 is a simple eigenvalue of A_0 . By simple I mean that its multiplicity is 1.

0 is a simple eigenvalue of A_0 .

Definition:

The *tiniest eigenvalue* of a matrix is a simple eigenvalue

that is *closer* to zero than any other eigenvalue.

Of course not every matrix has a tiniest eigenvalue. If the matrix is real and has a tiniest eigenvalue then the tiniest eigenvalue must be real. If it weren't real then its complex conjugate would be another eigenvalue the same distance from zero. We see that A_0 has a tiniest eigenvalue, and it's 0.

We know that the eigenvalues of a matrix are continuous functions of the matrix entries, so the eigenvalues of A_{δ} are continuous functions of δ . So there is some interval $[0, \nu]$ of reals such that if δ is in the interval then A_{δ} has a tiniest eigenvalue. The upper boundary ν is a small positive number. Several times during this discussion I will decrease the value of ν , shortening the interval. But I will always keep ν positive. During our discussion, we will assume without saying it that δ is in the interval.

So I can simply say that A_{δ} has a timest eigenvalue. I will call it λ . The eigenvalue λ is a continuous function of δ .

4.2 Eigenvectors with Eigenvalue λ

We define $M_{\delta} = A_{\delta} - \lambda I$. Since λ is a simple eigenvalue, the eigenvectors of A_{δ} with eigenvalue λ form a one dimensional subspace. That subspace is the kernel of the transformation M_{δ} . So the range of M_{δ} has dimension $\bar{N} - 1$. So M_{δ} has a nonzero $(\bar{N} - 1) \times (\bar{N} - 1)$ minor, and a nonzero cofactor. So $\operatorname{adj}(M_{\delta})$ is not the zero matrix.⁶

Select indices i and j such that the ij'th entry in $\operatorname{adj}(M_0)$ is nonzero. Remember those indices. So we see that the entry $\mathbf{e}_i(\operatorname{adj}(M_\delta))\mathbf{e}_j^\top$ is a continuous function of δ , and that it is nonzero when $\delta = 0$. If necessary, we now reduce the size of ν , keeping it positive, so that as long as δ is in the now shorter interval, the entry will be nonzero. The *i*'th row will be nonzero and so will the *j*'the column. (That is, they both have a nonzero entry.)

⁶By $\operatorname{adj}(M_{\delta})$ I mean the adjugate of M_{δ} .

Let the vector \mathbf{z} be the normalized *i*'th row of $\operatorname{adj}(M_{\delta})$ and let $\mathbf{\ddot{z}}^{\top}$ be the normalized *j*'th column. So \mathbf{z} and $\mathbf{\ddot{z}}^{\top}$ are both real unit vectors, and they are continuous functions of δ .

Now M_{δ} is singular, so we have $(\operatorname{adj}(M_{\delta})) M_{\delta} = |M_{\delta}| I = 0$. So the *i*'th row of $\operatorname{adj}(M_{\delta})$ is in the kernel of M_{δ} , and consequently it is a left eigenvector of A_{δ} with eigenvalue λ . So we see that \mathbf{z} is a left eigenvector of A_{δ} with eigenvalue λ . By the same argument beginning with $M_{\delta}(\mathrm{adj}(M_{\delta})) = |M_{\delta}|I = 0$, we see that $\mathbf{\ddot{z}}^{\top}$ is a right eigenvector of A_{δ} with eigenvalue λ .

Let's write \mathbf{z}_0 for the vector \mathbf{z} when $\delta = 0$. When $\delta = 0$, the eigenvalue λ is 0 and both \mathbf{z}_0 and $\boldsymbol{\eta}$ are left eigenvectors of A_0 with eigenvalue λ . Since λ is a simple eigenvalue, the space of such eigenvectors is one dimensional, so there is a scalar α such that $\eta = \alpha \mathbf{z}_0$. Since \mathbf{z} and \mathbf{e} are both real unit vectors, α must be either +1 or -1. We define the unit vector $\boldsymbol{\zeta} = \alpha \mathbf{z}$. Of course if $\delta = 0$ then $\boldsymbol{\zeta} = \alpha \, \mathbf{z}_0 = \boldsymbol{\eta}$.

So ζ is a real unit vector that is a continuous function of δ . It is a left eigenvector of A_{δ} with eigenvalue λ . If $\delta = 0$ then $\zeta = \eta$.

In exactly the same way, we define the column vector $\,\,\ddot{\zeta}^{\, op}$. It is a real unit column vector and it is a continuous function of δ . It is a right eigenvector of A_{δ} with eigenvalue λ . If $\delta = 0$ then $\ddot{\boldsymbol{\zeta}}^{ op} = \boldsymbol{\eta}^{ op}$

So $\zeta \ddot{\zeta}^{\top}$ is a real scalar that is a continuous function of δ . And it's 1 if $\delta = 0$. So if necessary we can reduce ν yet again and ensure that if δ is in the interval $[0,\nu]$ then $\zeta \ddot{\zeta}^{\top} > 0$. We define $\hat{\alpha} = (\boldsymbol{\zeta} \, \boldsymbol{\ddot{\zeta}}^{\top})^{-1}$

So $\hat{\alpha}$ is positive, and if $\delta = 0$ then $\hat{\alpha} = 1$.

4.3 Limit of Projected Eigenvector

We define the vector

 $\tau = \mathbf{\tilde{p}} \Psi$.

The real scalar $\zeta \eta^{\top}$ is a continuous function of δ , and it is 1 when $\delta = 0$. So if we reduce ν yet again we can ensure that $\zeta \eta^{\top}$ is positive for all δ in the interval $[0,\nu]$. We define the vector
$$\begin{split} \mathbf{v} &= \left(\boldsymbol{\zeta}\,\boldsymbol{\eta}^{\top}\right)^{-1}\boldsymbol{\zeta} \ . \\ \mathbf{v}\,\boldsymbol{\eta}^{\top} &= \ 1 \end{split}$$

And **v** is a left eigenvector of A_{δ} with eigenvalue λ . If $\delta = 0$ then $\mathbf{v} = \boldsymbol{\eta}$. Now from $Y_{\delta} = Y_0 + \delta F$, we have $A_{\delta} = A_0 + \delta \Psi^{\top} F \Psi$. Since $\eta \Psi^{\top} = \beta \mathbf{e}$, we have $\eta A_{\delta} = \delta \eta \Psi^{\top} F \Psi = \delta \beta \mathbf{e} F \Psi = \delta \beta \mathbf{p} \Psi = \delta \beta \tau$ and $A_{\delta} \boldsymbol{\eta}^{\top} = \delta \Psi^{\top} F \Psi \boldsymbol{\eta}^{\top} = \delta \beta \Psi^{\top} F \mathbf{e}^{\top} = \delta \beta \Psi^{\top} \tilde{\mathbf{p}}^{\top} = \delta \beta \boldsymbol{\tau}^{\top}.$

$$\boldsymbol{\eta} A_{\delta} = \delta \beta \boldsymbol{\tau} \qquad \text{and} \qquad A_{\delta} \boldsymbol{\eta}^{\top} = \delta \beta \boldsymbol{\tau}^{\top}$$
(8)

Then since $\mathbf{v}A_{\delta} = \lambda \mathbf{v}$, we have $\lambda = \lambda \mathbf{v} \boldsymbol{\eta}^{\top} = \mathbf{v}A_{\delta} \boldsymbol{\eta}^{\top} = \delta \beta (\mathbf{v} \boldsymbol{\tau}^{\top})$.

$$\lambda = \delta \beta \left(\mathbf{v} \, \boldsymbol{\tau}^{\top} \right) \tag{9}$$

We have $\eta \Psi^{\top} = \beta \mathbf{e}$. If we multiply by $\mathbf{\tilde{p}}^{\top}$ on the right, we obtain $\eta \tau^{\top} = \beta ,$

and this is positive. So if $\delta = 0$ then $\mathbf{v} = \boldsymbol{\eta}$, and $\mathbf{v} \boldsymbol{\tau}^{\top}$ is β , which is positive. So if we again appropriately reduce ν , we can ensure that $\mathbf{v} \boldsymbol{\tau}^{\top}$ is positive for any δ in the interval $[0, \nu]$. So $\mathbf{v} \boldsymbol{\tau}^{\top} > 0$ and $\beta > 0$. Therefore, (9) tells us that for any δ in the interval:

 λ is positive if δ is positive.

 λ is zero if δ is zero.

We define the vector $\boldsymbol{\omega}_{\delta} \;=\; \mathbf{v} - (\mathbf{v}\, \boldsymbol{ au}^{ op})^{-1} \boldsymbol{ au} \;.$ This vector is a continuous function of δ , since **v** is. We define a projection G onto the space of vectors orthogonal^{*} to η . $G = I - \boldsymbol{\eta}^\top \boldsymbol{\eta} .$ $G A_{\delta} = A_{\delta} - \boldsymbol{\eta}^{\top}(\boldsymbol{\eta} A_{\delta}) = A_{\delta} - \delta \beta \boldsymbol{\eta}^{\top} \boldsymbol{\tau}$ $\mathbf{v} G A_{\delta} = \mathbf{v} A_{\delta} - \delta \beta (\mathbf{v} \boldsymbol{\eta}^{\top}) \boldsymbol{\tau} = \lambda \mathbf{v} - \delta \beta \boldsymbol{\tau}$ Using (9) and the last equation gives us $\lambda \boldsymbol{\omega}_{\delta} = \lambda \mathbf{v} - \lambda \left(\mathbf{v} \,\boldsymbol{\tau}^{\top}\right)^{-1} \boldsymbol{\tau} = \lambda \mathbf{v} - \delta \beta \,\boldsymbol{\tau} = \mathbf{v} \, G \, A_{\delta} \,.$ Since $G \, \boldsymbol{\eta}^{\top} = 0$ and $B_{\delta} = A_{\delta} + \boldsymbol{\eta}^{\top} \boldsymbol{\eta}$, we have $G B_{\delta} = G \, A_{\delta}$ and $\mathbf{v} G B_{\delta} = \lambda \, \boldsymbol{\omega}_{\delta} \, .$

Now suppose $\delta > 0$. Then $\lambda > 0$, so the equation the end of the last paragraph can be written $\frac{1}{\lambda} \mathbf{v} G = \boldsymbol{\omega}_{\delta} B_{\delta}^{-1}$. By the definition of \mathbf{v} we have $(\boldsymbol{\zeta} \boldsymbol{\eta}^{\top}) \mathbf{v} = \boldsymbol{\zeta}$, so multiplying the last equation by $(\boldsymbol{\zeta} \boldsymbol{\eta}^{\top})$ gives $\frac{1}{\lambda} \boldsymbol{\zeta} G = (\boldsymbol{\zeta} \boldsymbol{\eta}^{\top}) \boldsymbol{\omega}_{\delta} B_{\delta}^{-1}$. We now let $\delta \to 0$. We see that $\boldsymbol{\omega}_{0} = \boldsymbol{\eta} - \beta^{-1} \boldsymbol{\tau}$.

$$\lim_{\delta \to 0} \frac{1}{\lambda} \zeta G = \omega_0 B_0^{-1} \tag{10}$$

4.4 Tidying up

Still assuming $\delta > 0$, we can show $\left(A_{\delta} + (1-\lambda)\,\hat{\alpha}\,\boldsymbol{\ddot{\zeta}}^{\top}\boldsymbol{\zeta}\right)\left(A_{\delta}^{-1} + (1-\frac{1}{\lambda})\,\hat{\alpha}\,\boldsymbol{\ddot{\zeta}}^{\top}\boldsymbol{\zeta}\right) = I$ simply by multiplying out and using these facts. $A_{\delta} A_{\delta}^{-1} = I \qquad A_{\delta} \ddot{\boldsymbol{\zeta}}^{\top} = \lambda \ddot{\boldsymbol{\zeta}}^{\top} \qquad \boldsymbol{\zeta} A_{\delta}^{-1} = \frac{1}{\lambda} \boldsymbol{\zeta}$ $\boldsymbol{\zeta} \, \boldsymbol{\ddot{\zeta}}^{ op} = \hat{\alpha}$ Therefore, we have this. $(A_{\delta}^{-1} + (1 - \frac{1}{\lambda})\hat{\alpha}\,\ddot{\boldsymbol{\zeta}}^{\top}\boldsymbol{\zeta}) = (A_{\delta} + (1 - \lambda)\hat{\alpha}\,\ddot{\boldsymbol{\zeta}}^{\top}\boldsymbol{\zeta})^{-1}$ $A_{\delta}^{-1} = (A_{\delta} + (1 - \lambda)\hat{\alpha}\,\ddot{\boldsymbol{\zeta}}^{\top}\boldsymbol{\zeta})^{-1} - \hat{\alpha}\,\ddot{\boldsymbol{\zeta}}^{\top}\boldsymbol{\zeta} + \frac{1}{\lambda}\hat{\alpha}\,\ddot{\boldsymbol{\zeta}}^{\top}\boldsymbol{\zeta}$ We multiply by G on the right and obtain this nice equation. $\begin{array}{l} A_{\delta}^{-1}G &= \left(A_{\delta} + (1-\lambda)\,\hat{\alpha}\,\ddot{\boldsymbol{\zeta}}^{\top}\boldsymbol{\zeta}\right)^{-1}G - \hat{\alpha}\,\ddot{\boldsymbol{\zeta}}^{\top}\boldsymbol{\zeta}\,G + \hat{\alpha}\,\ddot{\boldsymbol{\zeta}}^{\top}(\frac{1}{\lambda}\,\boldsymbol{\zeta}\,G)\\ \text{We now let}\quad\delta\to 0 \ . \quad \text{We have the following limits.}\\ \lambda\to 0 \quad \hat{\alpha}\to 1 \quad \boldsymbol{\zeta}^{\top}\to \boldsymbol{\eta}^{\top} \quad \boldsymbol{\zeta}\to \boldsymbol{\eta} \quad A_{\delta}\to A_{0} \end{array}$ and B_0 is nonsingular, so $(A_{\delta} + (1 - \lambda) \hat{\alpha} \, \boldsymbol{\zeta}^{\top} \boldsymbol{\zeta})^{-1} \to B_0^{-1}$. $(A_{\delta} + (1 - \lambda) \hat{\alpha} \, \boldsymbol{\zeta}^{\top} \boldsymbol{\zeta}) \to B_0$ From these limits and (10) we see that our nice equation becomes $\lim_{\delta \to 0} A_{\delta}^{-1} G = B_0^{-1} G + \boldsymbol{\eta}^\top \boldsymbol{\omega}_0 B_0^{-1} .$ Since η is happy, we have $\eta B_0 = \eta$ and $\eta = \eta B_0^{-1}$. $\omega_0 B_0^{-1} = \eta - \beta^{-1} \tau B_0^{-1}$ $\lim_{\delta \to 0} A_{\delta}^{-1} G = B_0^{-1} G + \eta^{\top} \eta - \beta^{-1} \eta^{\top} \tau B_0^{-1}$ We multiply on the right by $\Psi^{\top}D\mathbf{a}^{\top}$ and use $\eta \Psi^{\top}D\mathbf{a}^{\top} = \beta \mathbf{e} D\mathbf{a}^{\top} = 0$ and $G\Psi^{\top}D\mathbf{a}^{\top} = \Psi^{\top}D\mathbf{a}^{\top}$. $\lim_{\delta \to 0} A_{\delta}^{-1}\Psi^{\top}D\mathbf{a}^{\top} = B_{0}^{-1}\Psi^{\top}D\mathbf{a}^{\top} - \beta^{-1}\eta^{\top}\tau B_{0}^{-1}\Psi^{\top}D\mathbf{a}^{\top}$. Since $\mathbf{u}_{\delta}^{\top} = A_{\delta}^{-1}\Psi^{\top}D\mathbf{a}^{\top}$ and $\mathbf{u}^{\top} = B_{0}^{-1}\Psi^{\top}D\mathbf{a}^{\top}$, we $\lim_{\delta \to 0} \mathbf{u}_{\delta}^{\top} = \mathbf{u}^{\top} - \beta^{-1}\eta^{\top}(\tau \mathbf{u}^{\top})$. we have We define $\mathbf{u}_0^{\top} = \lim_{\delta \to 0} \mathbf{u}_{\delta}^{\top}$. Then we have $\mathbf{u}_0^{\top} = \mathbf{u}^{\top} - (\boldsymbol{\tau} \, \mathbf{u}^{\top}) \, \beta^{-1} \boldsymbol{\eta}^{\top} \; .$

5 Value Estimates

5.1 Value Estimate Notation

We said that the vector of value estimates was $\mathbf{\bar{v}}^{\top} = \Psi \mathbf{v}^{\top}$. But the average value $\mathbf{\tilde{p}} \mathbf{c}^{\top}$ is zero so we would like the average estimate $\mathbf{\tilde{p}} \mathbf{\bar{v}}^{\top}$ to be zero too. So to get better estimates we adjust the estimates by subtracting the average estimate from each estimate. So a better vector of estimates is $\mathbf{\ddot{v}}^{\top} = \mathbf{\bar{v}}^{\top} - (\mathbf{\tilde{p}} \mathbf{\bar{v}}^{\top}) \mathbf{e}^{\top} = (I - \mathbf{e}^{\top} \mathbf{\tilde{p}}) \mathbf{\bar{v}}^{\top}$. We now have $\mathbf{\tilde{p}} \mathbf{\ddot{v}}^{\top} = 0$. Note that $(\mathbf{\ddot{v}}_j - \mathbf{\ddot{v}}_i + a_i) = (\mathbf{\bar{v}}_j - \mathbf{\bar{v}}_i + a_i)$, so it makes no difference if TD(0) uses $\mathbf{\ddot{v}}^{\top}$ instead of $\mathbf{\bar{v}}^{\top}$.

The vector \mathbf{v} might be \mathbf{u} or \mathbf{u}_{δ} . Then we can write $\mathbf{\bar{u}}$, $\mathbf{\ddot{u}}$, $\mathbf{\bar{u}}_{\delta}$, and $\mathbf{\ddot{u}}_{\delta}$.

5.2 The Sad Terms

In our formulae for limit cash balance column vectors we have sad terms and happy terms. It is really only the sad terms that are important. To obtain our column vector of state value estimates, we multiply the limit column vector by Ψ on the left. So a term that is in the kernel has no effect at all. A happy term that is not in the kernel, when multiplied by Ψ , becomes a tidy vector, so it does change the estimates, but the change in each estimate is the same. The estimates of the values c_i become uniformly too high or too low. Our adjustment gets rid of such a change, but in many adaptation applications this uniform change makes no difference anyway. For example, what I call the choice value of transition $i \to j$ is this. $c_j - c_i + a_i$. In estimating choice value, the uniform changes cancel out.

If $\mathcal{H} \neq \mathcal{K}$ then our formula for \mathbf{u}_0^{\top} is this. $\mathbf{u}^{\top} - (\boldsymbol{\tau} \, \mathbf{u}^{\top}) \beta^{-1} \boldsymbol{\eta}^{\top}$ The second term is happy. $\mathbf{\bar{u}}_0^{\top} = \Psi \, \mathbf{u}_0^{\top} = \Psi \, \mathbf{u}^{\top} - (\boldsymbol{\tau} \, \mathbf{u}^{\top}) \beta^{-1} (\Psi \, \boldsymbol{\eta}^{\top}) = \mathbf{\bar{u}}^{\top} - (\mathbf{\tilde{p}} \, \Psi \, \mathbf{u}^{\top}) \beta^{-1} (\beta \, \mathbf{e}^{\top}) = \mathbf{\bar{u}}^{\top} - (\mathbf{\tilde{p}} \, \mathbf{\bar{u}}^{\top}) \mathbf{e}^{\top}$ And we see that $\bar{\mathbf{u}}_0$ is already adjusted. $\ddot{\mathbf{u}}_0 = \bar{\mathbf{u}}_0$. The happy term uniformly raises (or lowers) the estimates by just enough to make the average of the estimates zero.

5.3 What if $\Psi = I$?

From equation (1) we have $\tilde{\mathbf{p}} \mathbf{c}_{\delta}^{\top} = 0$ and $(I - (1 - \delta) P) \mathbf{c}_{\delta}^{\top} = \mathbf{a}^{\top} .$ $Y_{\delta} \mathbf{c}_{\delta}^{\top} = D \mathbf{a}^{\top} .$

Now assume N = N and $\Psi = I$.

Linear-TD(0) then becomes what Evolutionary Computation literature calls a simple bucket brigade on a Markov chain. In that case we have:

$$K = 0 \quad \boldsymbol{\eta} = N^{-\frac{1}{2}} \mathbf{e} \qquad H = \frac{1}{N} \mathbf{e}^{\mathsf{T}} \mathbf{e} \qquad \boldsymbol{\tau} = \tilde{\mathbf{p}} \qquad \beta = N^{-\frac{1}{2}} \qquad A_{\delta} = Y_{\delta}$$
$$A_{\delta} \mathbf{c}_{\delta}^{\mathsf{T}} = D \mathbf{a}^{\mathsf{T}}$$

If $\delta > 0$ then A_{δ} is nonsingular and $\mathbf{c}_{\delta}^{\top} = A_{\delta}^{-1} D \mathbf{a}^{\top}$. In other words, $\mathbf{u}_{\delta}^{\top} = \mathbf{c}_{\delta}^{\top}$.

Now suppose $\delta = 0$. We have: $Y \mathbf{c}^{\top} = D \mathbf{a}^{\top} \qquad B_0 = Y + \frac{1}{N} \mathbf{e}^{\top} \mathbf{e}$ Define $\chi = \frac{1}{N} \mathbf{e} \mathbf{c}^{\top}$. We now have: $B_0 \mathbf{c}^{\top} = D \mathbf{a}^{\top} + \chi \mathbf{e}^{\top} \qquad B_0 \mathbf{e}^{\top} = \mathbf{e}^{\top} \qquad B_0 (\mathbf{c}^{\top} - \chi \mathbf{e}^{\top}) = D \mathbf{a}^{\top}$ $\mathbf{c}^{\top} - \chi \mathbf{e}^{\top} = B_0^{-1} D \mathbf{a}^{\top} = \mathbf{u}^{\top}$ $\boldsymbol{\tau} \mathbf{u}^{\top} = \tilde{\mathbf{p}} \mathbf{u}^{\top} = -\chi \qquad \beta^{-1} \boldsymbol{\eta}^{\top} = \mathbf{e}^{\top}$ $\mathbf{u}_0^{\top} = \mathbf{u}^{\top} - (\boldsymbol{\tau} \mathbf{u}^{\top}) \beta^{-1} \boldsymbol{\eta}^{\top} = (\mathbf{c}^{\top} - \chi \mathbf{e}^{\top}) - (-\chi) \mathbf{e}^{\top} = \mathbf{c}^{\top}$ So we see that whether δ is zero or positive, we have:

If $\Psi = I$, then $\mathbf{u}_{\delta}^{\top} = \mathbf{c}_{\delta}^{\top}$.

The simple bucket brigade on a Markov chain converges to the correct values.⁷

5.4 False Payoffs in the Undiscounted Case

So we have $\mathbf{\ddot{v}}^{\top} = (I - \mathbf{e}^{\top} \mathbf{\tilde{p}}) \Psi \mathbf{v}^{\top}$. Note that if \mathbf{v} is happy then $\mathbf{\ddot{v}}^{\top} = 0$. Now let's look at the undiscounted case and suppose \mathbf{v} is the limit vector $\mathbf{u} + \mathbf{y}$, where \mathbf{y} is happy. So $\mathbf{\ddot{v}}^{\top}$ in the limit is $\mathbf{\ddot{u}}^{\top}$, where $\mathbf{\ddot{u}}^{\top} = (I - \mathbf{e}^{\top} \mathbf{\tilde{p}}) \Psi \mathbf{u}^{\top}$.

The vector $\mathbf{\ddot{u}}^{\top}$ is a vector of our estimates \ddot{u}_i of values c_i . The estimates can be very wrong. To call them approximations would be misleading. I call \ddot{u}_i the false value of i.

We have $\tilde{\mathbf{p}} \mathbf{\ddot{u}}^{\top} = 0$, and we define $\hat{\mathbf{a}}^{\top} = (I - P) \, \ddot{\mathbf{u}}^{\top} \, .$ I call \hat{a}_i the false payoff of i. We see that $\mathbf{\hat{a}}^{\top} = (I - P) (I - \mathbf{e}^{\top} \mathbf{\tilde{p}}) \Psi \mathbf{u}^{\top} = (I - P) \Psi \mathbf{u}^{\top}$, so we have $\mathbf{\hat{a}}^{\top} = (I - P) \Psi \mathbf{u}^{\top} .$ We define the matrix $\hat{P} = I - P + \mathbf{e}^{\top} \tilde{\mathbf{p}} .$ We note that \hat{P} is nonsingular.⁸ We note that $\tilde{\mathbf{p}}\mathbf{c}^{\top} = 0$ and $\mathbf{a}^{\top} = (I-P)\mathbf{c}^{\top}$. We have $\hat{P} \mathbf{c}^{\top} = \mathbf{a}^{\top}$ and $\mathbf{c}^{\top} = \hat{P}^{-1} \mathbf{a}^{\top} = \sum_{n=0}^{\infty} (P^n \mathbf{a}^{\top})$. We can think of $\hat{P}^{-1}\mathbf{a}^{\top}$ as an alternative definition of \mathbf{c}^{\top} . We also have $\hat{P}\mathbf{\ddot{u}}^{\top} = \mathbf{\hat{a}}^{\top}$, so $\ddot{\mathbf{u}}^{\top} = \hat{P}^{-1} \, \hat{\mathbf{a}}^{\top} = \sum_{n=0}^{\infty} (P^n \, \hat{\mathbf{a}}^{\top}) \; .$

⁷That statement was proved directly in [2].

⁸Suppose $\mathbf{v}\hat{P} = 0$ for some non-zero vector \mathbf{v} . Since $\hat{P}\mathbf{e}^{\top} = \mathbf{e}^{\top}$, we have $\mathbf{v}\mathbf{e}^{\top} = \mathbf{v}\hat{P}\mathbf{e}^{\top} = 0$. Therefore, $\mathbf{v}(I-P) = \mathbf{v}\hat{P} = 0$, so \mathbf{v} is an eigenvector of P with eigenvalue 1. So is $\tilde{\mathbf{p}}$, and the Frobenius Perron theorem tells us that the space of such eigenvectors is one dimensional. Therefore, $\mathbf{v} = \lambda \tilde{\mathbf{p}}$ for some scalar λ . Therefore, $\mathbf{v} = \lambda \mathbf{\tilde{p}} = \lambda \mathbf{\tilde{p}} \hat{P} = \mathbf{v} \hat{P} = 0$. Contradiction.

The false values $\mathbf{\ddot{u}}^{\top}$ are what the true values \mathbf{c}^{\top} would be if the excess payoffs \mathbf{a}^{\top} were the false payoffs $\hat{\mathbf{a}}^{\top}$. If you change the payoffs to the false payoffs, then the values become the false values.

This paper does not discuss adaptation. Adaptation is the process of changing the probabilities P in an attempt to increase \bar{m} . Often adaptation is on the basis of state values \mathbf{c}^{\top} . But when we base our adaptation on estimates $\mathbf{\ddot{u}}^{\top}$ given to us by TD(0) rather than on the true values \mathbf{c}^{\top} , it's as if TD(0) has changed the true payoffs \mathbf{a}^{\top} into false payoffs $\mathbf{\hat{a}}^{\top}$. This change is a problem inherent in all temporal difference methods.⁹ What makes this all non-trivial is that there is an interesting relation between the true

payoffs \mathbf{a}^{\top} and the false payoffs $\mathbf{\hat{a}}^{\top}$, which we now show. Since \mathbf{u} is sad, we have, $H \mathbf{u}^{\top} = 0$, so $\Psi^{\top} D \mathbf{a}^{\top} = (B_0) (B_0^{-1} \Psi^{\top} D \mathbf{a}^{\top}) = (\Psi^{\top} Y \Psi + H) (\mathbf{u}^{\top}) = \Psi^{\top} Y \Psi \mathbf{u}^{\top}$.

$$\Psi^{\top} D \mathbf{a}^{\top} = \Psi^{\top} Y \Psi \mathbf{u}^{\top}$$
(11)

 $\Psi^{\top} D \, \hat{\mathbf{a}}^{\top} = \Psi^{\top} D \left(I - P \right) \Psi \, \mathbf{u}^{\top} = \Psi^{\top} Y \, \Psi \, \mathbf{u}^{\top} = \Psi^{\top} D \, \mathbf{a}^{\top}$

$$\Psi^{\top} D \, \hat{\mathbf{a}}^{\top} = \Psi^{\top} D \, \mathbf{a}^{\top} \qquad \sum_{i} \tilde{p}_{i} \psi_{ij} \hat{a}_{i} = \sum_{i} \tilde{p}_{i} \psi_{ij} a_{i} \qquad (12)$$

I regard $\psi_{ij}a_i$ as the payoff received by feature j when we are in state i. So the average payoff received by feature j in each step is $\sum_{i} \tilde{p}_{i} \psi_{ij} a_{i}$. We see from (12) that in the change from true payoffs to false payoffs, the average is unchanged. In a sense, the change conserves payoff at each feature. The change is a problem, but the conservation is encouraging.

In a sense, $\hat{\mathbf{a}}$ is the only long vector that conserves payoff at each feature in the sense of equation (12). We show this as follows. Suppose \mathbf{z} is a long real vector such that

 $\Psi^{\top} D \mathbf{z}^{\top} = \Psi^{\top} D \mathbf{a}^{\top}$ And suppose $\tilde{\mathbf{p}} \mathbf{z}^{\top} = 0$, and suppose there is a short real vector \mathbf{v} such that $\mathbf{z}^{\top} = (I - P) \Psi \mathbf{v}^{\top}$. Write \mathbf{v} as the sum of sad vector \mathbf{x} and happy vector \mathbf{y} . We see that in the last equation, the \mathbf{y} drops out and we have $\mathbf{z}^{\top} = (I - P) \, \Psi \, \mathbf{x}^{\top} \; .$ $\Psi^{\top} D \mathbf{z}^{\top} = \Psi^{\top} D (I - P) \Psi \mathbf{x}^{\top} = \Psi^{\top} Y \Psi \mathbf{x}^{\top} = B_0 \mathbf{x}^{\top}$ By (11) we have $\Psi^{\top} D \mathbf{a}^{\top} = B_0 \mathbf{u}^{\top} .$ Since $\Psi^{\top} D \mathbf{z}^{\top} = \Psi^{\top} D \mathbf{a}^{\top}$, we have $B_0 \mathbf{x}^{\top} = B_0 \mathbf{u}^{\top}$, so $\mathbf{x}^{\top} = \mathbf{u}^{\top}$. We then have $\mathbf{z}^{\top} = (I - P) \Psi \mathbf{x}^{\top} = (I - P) \Psi \mathbf{u}^{\top} = \mathbf{\hat{a}}^{\top}$

Other Linear-TD(0) Methods 6

6.1 Using \mathbf{m}^{\top} Instead of \mathbf{a}^{\top}

What if the method uses \mathbf{m}^{\top} instead of \mathbf{a}^{\top} ? Then the average step equation (4) becomes $(\mathbf{v}^{(n+1)})^{\top} = (I - \varepsilon \Psi^{\top} Y_{\delta} \Psi) (\mathbf{v}^{(n)})^{\top} + \varepsilon \Psi^{\top} D \mathbf{m}^{\top}$.

Suppose $\delta > 0$.

Then our convergence proof still works, but with \mathbf{m}^{\top} place of \mathbf{a}^{\top} . We write $(\mathbf{v}^{(n)})^{\top} = (\mathbf{x}^{(n)})^{\top} + \mathbf{y}^{\top}$, where $(\mathbf{x}^{(n)})^{\top}$ is the antikernel part and \mathbf{y}^{\top} is the kernel part. $\lim_{n\to\infty} (\mathbf{x}^{(n)})^{\top} = A_{\delta}^{-1} \Psi^{\top} D \mathbf{m}^{\top} = \mathbf{u}_{\delta}^{\top} + \bar{m} A_{\delta}^{-1} \boldsymbol{\tau}^{\top}$ What is this awkward term $\bar{m} A_{\delta}^{-1} \boldsymbol{\tau}^{\top}$?

If **y** is a kernel vector, then $\mathbf{y} A_{\delta} = \mathbf{y}$ and $\mathbf{y} A_{\delta}^{-1} = \mathbf{y}$. Therefore we have $\mathbf{y}(\bar{m}A_{\delta}^{-1}\boldsymbol{\tau}^{\top}) = \bar{m}(\mathbf{y}A_{\delta}^{-1})(\boldsymbol{\tau}^{\top}) = \bar{m}(\mathbf{y})(\Psi^{\top}\tilde{\mathbf{p}}^{\top}) = 0$, so the awkward term is an antikernel vector. If $\mathcal{H} = \mathcal{K}$, then the awkward term is sad.

If $\mathcal{H} \neq \mathcal{K}$ then we can use equations in section 4. Equation (8) tells us that the awkward term is this. $\bar{m} A_{\delta}^{-1} \boldsymbol{\tau}^{\top} = \delta^{-1} \beta^{-1} \bar{m} \boldsymbol{\eta}^{\top}$. In this case the awkward term is happy.

When calculating the vector of estimated values we multiply on the left by Ψ and the happy awkward term $\delta^{-1} \bar{m} \mathbf{e}^{\top}$. Since this is tidy, adjustment will get rid of it. But if δ is small the tidy term is becomes huge, and this can cause calculation problems. We now look at the even worse case when $\delta = 0$.

 $^{^{9}}$ Or if you prefer, the change is one way of viewing a problem that is inherent in all temporal difference methods.

Now suppose $\delta = 0$.

Most of our previous convergence proof works with \mathbf{m}^{\top} in place of \mathbf{a}^{\top} , but one part doesn't. We multiplied the average step equation by H and obtained $H(\mathbf{v}^{(n+1)})^{\top} = H(\mathbf{v}^{(n)})^{\top}$. We no longer get this. On the other hand, we do get $K(\mathbf{v}^{(n+1)})^{\top} = K(\mathbf{v}^{(n)})^{\top}$. So if $\mathcal{H} = \mathcal{K}$ then we do get $H(\mathbf{v}^{(n+1)})^{\top} = H(\mathbf{v}^{(n)})^{\top}$, and our proof works with \mathbf{m}^{\top} in place of \mathbf{a}^{\top} . We obtain $\lim_{n\to\infty} (\mathbf{v}^{(n)})^{\top} = B_0^{-1} \Psi^{\top} D \mathbf{m}^{\top} = \mathbf{u}^{\top} + \bar{m} B_0^{-1} \boldsymbol{\tau}^{\top} .$

But if $\delta = 0$ and $\mathcal{H} \neq \mathcal{K}$ then our convergence proof doesn't work. My guess is that in this case the sequence doesn't converge unless $\bar{m} = 0$.

TD(0)-future Methods 6.2

We define $\mathbf{b}^{\top} = P \mathbf{a}^{\top}$. Then we have $\sum_{w} F_{iw} b_{i} = \tilde{p}_{i} b_{i} = \tilde{p}_{i} \sum_{w} P_{iw} a_{w} = \sum_{w} F_{iw} a_{w}$, and therefore, $\varepsilon \sum_{iw} F_{iw} b_{i} \psi_{ik} = \varepsilon \sum_{iw} F_{iw} a_{w} \psi_{ik}$.

$$\varepsilon \sum_{iw} F_{iw}(b_i - \bar{v}_i + (1 - \delta) \, \bar{v}_w) \, \psi_{ik} = \varepsilon \sum_{iw} F_{iw}(a_w - \bar{v}_i + (1 - \delta) \, \bar{v}_w) \, \psi_{ik} \quad . \tag{13}$$

There is a class of TD(0) methods that I call TD(0)-future methods. In these methods, if the state transition is $i \to w$, the increment to v_k is $\varepsilon (a_w - \bar{v}_i + (1 - \delta) \bar{v}_w) \psi_{ik}$ rather than formula (2).

The TD(0)-future average change in v_k then is the right side of (13) rather than (3). Let's use the left side of (13). That's just (3) with each a_i replaced by b_i . So as far as average steps go, TD(0)-future behaves just like ordinary linear-TD(0) behaves except that the excess payoffs are \mathbf{b}^{\top} rather than \mathbf{a}^{\top} . All our convergence arguments also work for TD(0)-future, and the limit formulae are the same except that \mathbf{a}^{\top} is replaced by \mathbf{b}^{\top} , and \mathbf{m}^{\top} is replaced by $\mathbf{b}^{\top} + \bar{m} \mathbf{e}^{\top}$, which is $P \mathbf{m}^{\top}$.

Instead of limit vectors $\mathbf{u}_{\delta}^{\top}$ and \mathbf{u}^{\top} , we have $\mathbf{z}_{\delta}^{\top} = A_{\delta}^{-1} \Psi^{\top} D \mathbf{b}^{\top}$ an

and $\mathbf{z}^{\top} = B_0^{-1} \Psi^{\top} D \mathbf{b}^{\top}$.

Our cash balance vector \mathbf{v} might be \mathbf{z} or \mathbf{z}_{δ} , so we can write $\bar{\mathbf{z}}$, $\ddot{\mathbf{z}}$, $\bar{\mathbf{z}}_{\delta}$, and $\ddot{\mathbf{z}}_{\delta}$.

The TD(0)-future method is not trying to estimate the values $\mathbf{c}_{\delta}^{\top} = \sum_{n=0}^{\infty} ((1-\delta)^n P^n \mathbf{a}^{\top})$. It's trying to estimate the future values $\bar{\mathbf{w}}_{\delta}^{\top} = \sum_{n=0}^{\infty} ((1-\delta)^n P^n \mathbf{b}^{\top})$. We see that $\mathbf{c}_{\delta}^{\top} = \mathbf{a}^{\top} + (1-\delta) \bar{\mathbf{w}}_{\delta}^{\top}$. So we would like our new estimates to be related as follows.

 $\ddot{\mathbf{u}}_{\delta}^{\top} = \mathbf{a}^{\top} + (1 - \delta) \ddot{\mathbf{z}}_{\delta}^{\top}$ $\ddot{\mathbf{u}}^{ op} = \mathbf{a}^{ op} + \ddot{\mathbf{z}}^{ op}$. and

We shall see that if $\mathbf{a} \in \operatorname{Ran}(\Psi^{\top})$ then we do have that relationship. Suppose $\mathbf{a} \in \operatorname{Ran}(\Psi^{\top})$ and $\delta > 0$.

Then there is a kernel vector \mathbf{x} and an antikernel vector \mathbf{y} such that

$$\begin{split} \Psi(\mathbf{x}^{\top} + \mathbf{y}^{\top}) &= \mathbf{a}^{\top} . & \text{And so we have} \quad \Psi \mathbf{y}^{\top} = \mathbf{a}^{\top} . \\ A_{\delta} \mathbf{y}^{\top} &= (\Psi^{\top} Y_{\delta} \Psi + K) \mathbf{y}^{\top} = \Psi^{\top} Y_{\delta} \mathbf{a}^{\top} = \Psi^{\top} D \left(I - (1 - \delta) P \right) \mathbf{a}^{\top} = \Psi^{\top} D \mathbf{a}^{\top} - (1 - \delta) \Psi^{\top} D \mathbf{b}^{\top} \\ \text{We multiply by } A_{\delta}^{-1} \text{ on the left.} \quad \mathbf{y}^{\top} = \mathbf{u}_{\delta}^{\top} - (1 - \delta) \mathbf{z}_{\delta}^{\top} \\ \text{Now we multiply by } \Psi \text{ on the left.} \quad \mathbf{a}_{\tau}^{\top} = \mathbf{u}_{\delta}^{\top} - (1 - \delta) \mathbf{z}_{\delta}^{\top} \end{split}$$
So we see that the vector $\mathbf{a}^{\top} + (1 - \delta) \mathbf{\ddot{z}}_{\delta}^{\top} - \mathbf{\ddot{u}}_{\delta}^{\top}$ is tidy.

Multiplying that tidy vector by $\tilde{\mathbf{p}}$ on the left gives zero, so the tidy vector itself is zero, and we have

If $\mathbf{a} \in \operatorname{Ran}(\Psi^{\top})$ and $\delta > 0$, then $\mathbf{\ddot{u}}_{\delta}^{\top} = \mathbf{a}^{\top} + (1 - \delta) \mathbf{\ddot{z}}_{\delta}^{\top}$.

Suppose $\mathbf{a} \in \operatorname{Ran}(\Psi^{\top})$.

Then there is a sad vector \mathbf{x} and a happy vector \mathbf{y} such that $\Psi(\mathbf{x}^{\top} + \mathbf{y}^{\top}) = \mathbf{a}^{\top} .$ We multiply by (I - P). $(I-P)\Psi\mathbf{x}^{\top} + (I-P)\Psi\mathbf{y}^{\top} = (I-P)\mathbf{a}^{\top}$ The vector $\Psi \mathbf{y}^{\top}$ is tidy, so the second term is zero, and we have We multiply on the left by $\Psi^{\top}D$. Since $H\mathbf{v}^{\top} = 0$ this 1 $(I-P)\Psi\mathbf{x}^{\top} = \mathbf{a}^{\top} - \mathbf{b}^{\top}.$ $\Psi^{\top} Y \Psi \mathbf{x}^{\top} = \Psi^{\top} D (\mathbf{a}^{\top} - \mathbf{b}^{\top}) .$ Since $H \mathbf{x}^{\top} = 0$, this becomes
$$\begin{split} \Psi & I \Psi \mathbf{x} - \Psi D (\mathbf{a}^{\top} - \mathbf{b}^{\top}) \\ B_0 \mathbf{x}^{\top} &= \Psi^{\top} D (\mathbf{a}^{\top} - \mathbf{b}^{\top}) \\ \mathbf{x}^{\top} &= B_0^{-1} \Psi^{\top} D (\mathbf{a}^{\top} - \mathbf{b}^{\top}) = \mathbf{u}^{\top} - \mathbf{z}^{\top} . \end{split}$$
 $\begin{aligned} & (\mathbf{x}^\top + \mathbf{y}^\top) + \mathbf{z}^\top - \mathbf{u}^\top = \mathbf{y}^\top \\ & \mathbf{a}^\top + \mathbf{\bar{z}}^\top - \mathbf{\bar{u}}^\top = \Psi \mathbf{y}^\top \end{aligned}$ Now multiply by Ψ on the left. Since $\Psi \mathbf{y}^{\top}$ is a tidy vector, we see that the vector $\mathbf{a}^{\top} + \mathbf{\ddot{z}}^{\top} - \mathbf{\ddot{u}}^{\top}$ is also tidy. Multiplying that tidy vector by $\tilde{\mathbf{p}}$ on the left gives zero, so the tidy vector itself is zero, and we have

If $\mathbf{a} \in \operatorname{Ran}(\Psi^{\top})$ then $\mathbf{\ddot{u}}^{\top} = \mathbf{a}^{\top} + \mathbf{\ddot{z}}^{\top}$.

6.3 Ensuring $\mathcal{H} \neq \mathcal{K}$ and $\mathbf{a} \in \mathbf{Ran}(\Psi^{\top})$

If one of the basis functions simply returns the number 1, that is, if there is a k such that $\psi_{ik} = 1$ for all i, then $\mathbf{e} \in \operatorname{Ran}(\Psi^{\top})$, and $\mathcal{H} \neq \mathcal{K}$. The payoff m_i of the current state i is available to the system, so we can arrange that one of the basis functions simply returns that payoff. Then there is a k such that $\psi_{ik} = m_i$ for all i. Then $\mathbf{m} \in \operatorname{Ran}(\Psi^{\top})$. If we also have $\mathbf{e} \in \operatorname{Ran}(\Psi^{\top})$, then $\mathbf{a} \in \operatorname{Ran}(\Psi^{\top})$.

7 What's New Here?

The proof of average step convergence of undiscounted linear-TD(0) is new. Though its basic idea is the same as the proof of the discounted case, the lack of discount necessitates an added trick to show convergence. The resulting limit formula is of course simpler than in the discounted case. It relates to true values undistorted by discounts, and when used in adaptation it facilitates analysis, though we do not here analyze the adaptation.

Another approach to the undiscounted case is to take the discounted limit formula and let $\delta \to 0$. We do that here and that is new. The two approaches yield different formulae, but the difference makes sense.

There is a large body of Reinforcement Learning work on linear-TD(0), but mostly on the *discounted* version. And there is a large body of Evolutionary Computation work on the simple bucket brigade, which is a special case of *undiscounted* linear-TD(0). By showing convergence of undiscounted linear-TD(0), the work here plugs a gap in our knowledge and brings these two bodies of work together.

References

- Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction, Second edition. MIT Press, Cambridge, Mass., 2018.
- [2] T. H. Westerdale. Quasimorphisms or queasymorphisms? Modeling finite automaton environments. In Gregory J. E. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 128–147, San Mateo, CA, 1991. Morgan Kaufmann.
- [3] Tom Westerdale. Learning Resembles Evolution the Markov Case. pages 1–12, 2023. unpublished.