Average Step Convergence of Undiscounted Linear-TD(0)

Tom Westerdale

January 5, 2025

Abstract

The temporal difference method linear-TD(0) makes successive probabilistic adjustments in its estimates of state values. 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. A step size parameter governs the step size, and that parameter is constant throughout the sequence. The 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.

Key Words: reinforcement learning, temporal difference, discount, average step convergence, value estimation, bucket brigade.

1 Introduction

Linear-TD(0) methods are temporal difference methods that are used in adaptive systems to provide estimates of state values. In each time step, the adaptive system takes an action, the environment changes its state, and payoff is received by the system. Payoff[1] depends on state. Each state has a *value*, and the value of the current state is a measure of how much payoff we expect in the future.[2]

A linear-TD(0) method iteratively refines its estimates of state values. In their book, Sutton and Barto outline a proof that linear-TD(0) converges, provided the future payoffs are discounted [8, pp. 206-7]. What this paper does is extend their proof to the undiscounted case.

By discounting we mean the following. 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 [8] usually writes γ where we write $1-\delta$. Writing $1-\delta$ makes the tricky parts of the proofs here easier to understand. We will not use the symbol γ in this paper. Following everyday practice, I call δ the discount.[3]

Linear-TD(0) methods are key temporal difference methods in the field of Reinforcement Learning. A proper discussion of them and the context in which they arose is in the Reinforcement Learning book by Sutton and Barto [8]. Temporal difference methods have a long history in Adaptive Systems research. Arthur Samuel's Checkers Player used an early temporal difference method [6]. John Holland's *bucket brigade* [4] has been widely used in Evolutionary Computation, and its pros and cons have long been debated [10]. We proved average step convergence of the *simple bucket brigade* on a Markov chain [11]. This simple bucket brigade captures the essence of temporal difference thinking. Other bucket brigades can be thought of as elaborations of the simple bucket brigade.

[1] Terminology 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.[8] What we call value is what many Evolutionary Adaptation researchers call fitness, yet the term fitness is often used to mean reproductive rate.

[2] In Reinforcement Learning, payoff can be any quantity that depends on state, and the value of the current state can be thought of as a prediction of some function of future payoffs, so temporal difference methods are fairly general prediction methods. (See [8] for a proper discssion.) But this paper discusses only linear-TD(0) convergence, and in that context, the value of the current state is simply a prediction of the sum of future payoffs or excess payoffs, as we shall describe formally in sections 2 and 6. (See equation (1).)

[3] I will not use the phrase "discount rate".

A key event in the development of temporal difference methods was the appearance in 1981 of the seminal paper by Sutton and Barto [7]. It outlines how to adjust synapse weights on a model neuron in such a way that the neuron becomes a predictor. Sutton and Barto had made a minor but brilliant change to the Hebb synapse reinforcement rule [3] [7]. The result was a bucket brigade generalization in which the synapses correspond to Samuel's checkerboard features.[4] Sutton and Barto later formalized their insight in what they now call linear-TD(0), and they used the term "temporal difference" to describe all these approaches.

Discounting is rare in Evolutionary Computation. The simple bucket brigade is undiscounted. It is a special case of linear-TD(0), but Sutton and Barto's linear-TD(0) convergence proof works only when linear-TD(0) is discounted. Here we extend their proof to undiscounted linear-TD(0), and that makes our bucket brigade convergence proof redundant.[5]

Linear-TD(0) operates in roughly the following way. (We will give the details in subsection 2.2.) Like Samuel, linear-TD(0) examines features of the current state, the current state of the checkerboard in Samuel's case. Samuel used features like center control, back row control, and advancement. So consider a system that uses linear-TD(0). Let $\hat{N}$ be the number of different features that the system uses. We can number the features from 1 to $\hat{N}$. Each feature has a value, which is determined by the state, and which changes as the state changes. If the current state is i , then the vector of current feature values is $\langle \psi_{i1}, \psi_{i2}, \psi_{i3}, ..., \psi_{i\hat{N}} \rangle$. The number ψ_{ik} is the value that feature k has whenever the state is i .

The system holds a vector of adjustable *weights* $\langle v_1, v_2, v_3, ..., v_{\hat{N}} \rangle$, one weight for each of the features. To obtain an estimate $\bar{v}_i$ of the value of state i , the system uses the formula

$$\bar{v}_i \;=\; \psi_{i1}v_1 + \psi_{i2}v_2 + \psi_{i3}v_3 + \cdots + \psi_{i\hat{N}}v_{\hat{N}} \; .$$

Linear-TD(0) iteratively adjusts the weights $v_1, v_2, v_3, ..., v_{\hat{N}}$ in successive steps in an attempt to improve the estimates of the state values. The basic temporal difference idea is that when the state transition is $i \to j$, the adjustment to each weight v_k is $\varepsilon\,(\bar{v}_j - \bar{v}_i)\,\psi_{ik}$. The small constant ε is the *step size parameter.* The successive adjustments give us a sequence of weights vectors, and it is convergence of such a sequence that we need to show.

But there is an added detail. Linear-TD(0) multiplies each received payoff by ε and adds it into the weights. We will give all the details in subsection 2.2. The payoffs keep disturbing the weights, and the actual sequence of weights vectors keeps dancing around. It doesn't converge in the simple sense of convergence. The convergence proofs here and in Sutton and Barto's book are of a weak convergence that I call *average step convergence.* It is convergence of a notional sequence of weights vectors that we create by taking successive average steps. In each time unit we determine the average of all the weight adjustments that could occur in the next step. We then take that average step. And then in the following time unit we take another average step, and so on. The resulting sequence converges. The proofs assume that the states and state transitions form a Markov chain[6], so the averages are well defined. At the end of subsection 2.2 we give a proper definition of the averages.[7]

Some people take a different approach to this convergence problem.[1] Instead of having a constant ε, they reduce ε in each time step, letting it approach 0. This gradually reduces the disturbance in the weights. If they reduce ε fast enough, they get *convergence with probability* 1 of the actual sequence of weights vectors. The approach says how fast that reduction has to be. But reducing ε in this way has disadvantages, disadvantages that are outside the scope of this paper.[8] In this paper, we keep ε constant and the convergence that we prove is average step convergence.

Average step convergence gives us a conceptual limit vector, but the actual sequence of weights vectors keeps dancing. It would be nice if we could show that most of the time it is dancing near the conceptual limit, and that the smaller our constant ε is, the closer to the limit the sequence dances. Formally, this is *convergence in probability* as ε goes to zero.[9] This has been shown for the simple bucket brigade.[12] I believe such convergence holds for linear-TD(0) in general, but that has not been proved, and this paper does not discuss it. Our discussion here is limited to average step convergence.

In this paper, we extend Sutton and Barto's average step convergence proof so that it covers the undis-

[4] The synaptic inputs are the feature values.

[5] In subsection 5.2 we discuss the simple bucket brigade as a special case of linear-TD(0).

[6] strongly connected finite-state Markov chain

[7] Of course we can't really take an average step, but we pretend that we can. Convergence terminology varies across the literature. I call this average step convergence because at least that's clear. It's convergence of the sequence formed by taking average steps.

[8] For example, suppose we are in a continuing task in which payoff keeps arriving and the whole process conceptually goes on forever. And suppose the probabilities are being altered very slowly, perhaps by adaptation. So the state values are slowly changing, and it is the job of linear-TD(0) to track the changing values. (See [8] and [14] for discussion of this "actor-critic" scenario.) If ε approaches 0, the tracking comes to a halt.

[9] The convergence in probability does not say that we are reducing ε during a sequence. It says that a sequence with an ε that's closer to zero is a sequence that dances closer to the limit.

counted case. For comparison, we also repeat the shorter discounted case proof, using our notation, and we exhibit the limit formulae for both cases.[10]

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 state value estimates, which I call *false values* simply because they are not the true values. They are very useful, yet they can be very wrong and biased. We write them in terms of what I call *false payoffs* and prove a key relation between true and false payoffs. (equation(13)) The biases in temporal difference value estimates have long bedeviled our work in Evolutionary Computation, though that work has been exclusively with bucket brigades that are special cases of linear-TD(0).[11] The key relation proved here is general and gives us a handle on these biases.

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

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[12] 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 $i \to j$.
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 probabilty the chain is in state i is $\tilde{p}_i$. Every $\tilde{p}_i$ is positive.[13]

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, the system receives 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, $\bar{m} = \tilde{\mathbf{p}}\,\mathbf{m}^\top$.
A state's excess payoff is the amount by which its payoff exceeds the average payoff.
The vector $\mathbf{a}$ is the vector of *excess payoffs.* That is, $\mathbf{a} = \mathbf{m} - \bar{m}\,\mathbf{e}$.
The average excess payoff is of course zero. $\tilde{\mathbf{p}}\,\mathbf{a}^\top = 0$.

The column vector $\mathbf{c}^\top$ is the column vector of *state values.* Its definition is the Cesaro sum
$\mathbf{c}^\top = \sum\!\!\!\!\!{\scriptstyle\mathrm{C}}_{n=0}^{\infty}(P^n\,\mathbf{a}^\top)$. The sum converges in the Cesaro sense.[14]

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

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

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 $\mathbf{c}_0^\top = \mathbf{c}^\top$.

[10] The discounted case proof in subsection 3.3 uses Geršgorin's Theorem. Geršgorin's Theorem itself is not quite suitable for the undiscounted case, but we can use its proof [9, page 4] if we suitably modify it. The modified proof is the proof of Lemma 1.

[11] For example, see [10].

[12] By *strongly connected* I mean that for any ordered pair of states $\langle i, j\rangle$, there is a sequence of legal transitions that takes the chain from state i to state j .

[13] Since the chain is strongly connected, the absolute state probabilities are well defined.[5] That they are all positive follows directly from the Perron-Frobenius theorem.[2]

[14] For proof in this notation that it converges, see [13]. I write $\sum\!\!\!\!\!{\scriptstyle\mathrm{C}}$ for Cesaro sum. 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.

2.2 Linear-TD(0) and its Average Step

We can now describe linear-TD(0). First we repeat more carefully what we said in the introduction about the feature values ψ_{ik} and value estimates $\bar{v}_i$.

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

To estimate the value of the current state, the system uses a vector $\mathbf{v} = \langle v_1, v_2, v_3, ...v_{\hat{N}} \rangle$ of adjustable real parameters that we call *weights.*[15] The estimate $\bar{v}_i$ of the value of state i is $\bar{v}_i = \sum_k \psi_{ik} v_k$. The column vector $\bar{\mathbf{v}}^\top$ of the estimates of the state values is then given by $\bar{\mathbf{v}}^\top = \Psi \, \mathbf{v}^\top$.

I think of $\hat{N}$ as being smaller than N , though it doesn't have to be. The vectors $\mathbf{e}$, $\tilde{\mathbf{p}}$, $\bar{\mathbf{v}}$, $\mathbf{a}$, 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 $\hat{N}$ 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 $\hat{\mathbf{e}}_k$ to be the short vector whose k'th entry is 1 and whose other entries are 0.

We will have big square $N \times N$ matrices and small square $\hat{N} \times \hat{N}$ matrices.
The matrix I is the $N \times N$ identity matrix.
The matrix ${\scriptstyle I}$ is the $\hat{N} \times \hat{N}$ 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 weights. The version of linear-TD(0) that we will concentrate on adjusts the weights in the following way.
If the state transition is $i \to j$, then each weight v_k is incremented by the amount

$$\varepsilon \, (a_i - \bar{v}_i + (1-\delta) \, \bar{v}_j) \, \psi_{ik} \quad . \tag{2}$$

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

(Sutton and Barto frequently use m_i in place of a_i in the increment formula (2). This adds the extra amount $\bar{m} \, \varepsilon \, \psi_{ik}$ to the increment formula. In the undiscounted case, these extra amounts accumulate and generally prevent convergence, though the discounted case still converges. We discuss all this in subsection 6.1. The main convergence proofs in this paper will use increment formula (2) for both the undiscounted case and the discounted case.[16])

So given that the current vector of weights is $\mathbf{v}$, the average change in v_k is

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

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

$$Y_\delta \;=\; D \, (I - (1-\delta) \, P) \quad .$$

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

$$\varepsilon \, \Psi^\top D \, \mathbf{a}^\top - \varepsilon \, \Psi^\top Y_\delta \Psi \, \mathbf{v}^\top \quad . \tag{4}$$

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

$$(\mathbf{v}^{(n+1)})^\top = \; ({\scriptstyle I} - \varepsilon \, \Psi^\top Y_\delta \Psi) \, (\mathbf{v}^{(n)})^\top + \varepsilon \, \Psi^\top D \, \mathbf{a}^\top \quad . \tag{5}$$

[15] The weights are called cash balances in the bucket brigade.

[16] Another thing that Sutton and Barto frequently do is to use a_j in place of a_i in the increment formula, giving an increment of $\varepsilon \, (a_j - \bar{v}_i + (1-\delta) \, \bar{v}_j) \, \psi_{ik}$. That formulation is entirely equivalent to (2). We discuss the equivalence in subsection 6.2. I think the formulation used here makes the proofs slightly cleaner.

We ask whether this sequence converges, and if so, what it converges to. I call this convergence *average step convergence.*[17] Average step convergence has already been shown for $\delta > 0$. (See [8, 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.

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 $\mathbf{z}$ 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. $(\mathrm{Re}(\lambda))\mathbf{z}\,\mathbf{z}^* = \mathrm{Re}(\lambda\,\mathbf{z}\,\mathbf{z}^*) = \mathrm{Re}(\mathbf{z}A\,\mathbf{z}^*) = \mathbf{x}A\,\mathbf{x}^\top + \mathbf{y}A\,\mathbf{y}^\top$.
This is a positive real since both terms are non-negative and at least one is positive. So $\mathrm{Re}(\lambda) > 0$.
■

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

3.2 Undiscounted Y is Almost Positive Definite.

Notation: In this subsection all vectors are long vectors.

We call call an $N \times N$ real matrix A *nice* just if it has these four nice properties.
(1) Every diagonal entry of A is positive.
(2) Every off-diagonal entry of A is either negative or zero.
(3) For every legal transition $i \to j$ between two different states, the entry A_{ij} is negative.
(4) $\mathbf{e}A = 0$ and $A\mathbf{e}^\top = 0$.

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.

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

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

[17] That's not what it's called in the literature, but I find the terminology in the literature confusing.

Lemma 1
If for distinct states i *and* k *we have*
$z_i = 1$ *and* $S_{ik} \neq 0$ *, then :*
(1) $\lambda \geq 0$,
(2) $\lambda = 0 \implies z_k = 1$.

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 $\sum_j S_{ij} = \mathbf{e}_i S\,\mathbf{e}^\top = 0$, so

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

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,

$$S_{ii} - \lambda = \sum_{j\neq i} (\text{-}S_{ij})\, z_j \tag{7}$$

We have $\left|\sum_{j\neq i} (\text{-}S_{ij})\, z_j\right| \leq \sum_{j\neq i} (\text{-}S_{ij})\,|z_j| \leq \sum_{j\neq i} (\text{-}S_{ij})$. By (6) and (7) 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 (6) and (7) are equal, so $\sum_{j\neq i} (\text{-}S_{ij})(1 - z_j) = 0$, and $\sum_{j\neq i} (\text{-}S_{ij})(1 - x_j) = 0$. Since $x_j \leq |z_j| \leq 1$, every term in the last sum is non-negative. Hence every term is zero. In particular, $(\text{-}S_{ik})(1 - x_k) = 0$, so $x_k = 1$. This and $|z_k| \leq 1$ give us $z_k = 1$.
■

Let's call a state i green if $z_i = 1$, and let's call i red if $z_i \neq 1$. By $\mathbf{z}$ property (2), there is at least one green state.

By strong connectivity, there is from every state at least one legal transition to a different state.

Since S is nice, if $i \to k$ is a legal transition from state i to a different state k then $S_{ik} \neq 0$.

Lemma 1 has two conclusions.

Conclusion (1) tells us that if a state i is green and $i \to k$ is a legal transition from it to a different state, then $\lambda \geq 0$. Well in fact there does exist a green state and a legal transition from it to a different state, so conclusion (1) tells us the following simple fact.
$\lambda \geq 0$

Conclusion (2) deals with the rather special situation that obtains when $\lambda = 0$. In that case, if state i is green and $i \to k$ is a legal transition from i to a different state k , then state k is green too. So any state that a green state is connected to by a transition is also green. It follows that since the chain is strongly connected, every state is green, and $\mathbf{z} = \mathbf{e}$. So conclusion (2) tells us the following.
$\lambda = 0 \implies \mathbf{z} = \mathbf{e}$

Conversely, suppose $\mathbf{z} = \mathbf{e}$. Then $\lambda\mathbf{z} = \mathbf{z}S = \mathbf{e}S = 0$, so $\lambda = 0$.

So we have these results.
$\lambda = 0$ if and only if $\mathbf{z} = \mathbf{e}$.
$\lambda = 0$ if and only if $\mathbf{y}$ 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 $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \ldots., \mathbf{z}_N$ 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. Then λ_1 is zero, but every other eigenvalue λ_k is positive.

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.
Each $\mathbf{z}_k \mathbf{z}_k^\top$ is a positive real since each $\mathbf{z}_k$ is a real nonzero vector.
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 $\mathbf{x} = b_1 \mathbf{z}_1$, 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 have nice properties (1), (2), and (3) in the definition of nice matrix.
$\mathbf{e}\, S = 2\, \delta\, \tilde{\mathbf{p}}$ $\quad$ $S\, \mathbf{e}^\top = 2\, \delta\, \tilde{\mathbf{p}}^\top$

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 .
$\sum_j S_{ij} = \mathbf{e}_i S\, \mathbf{e}^\top = \mathbf{e}_i (2\delta\, \tilde{\mathbf{p}}^\top) = 2\delta\, \tilde{p}_i > 0$
$S_{ii} > \sum_{j \neq i} (-S_{ij})$ (The sum is over all j that are not equal to i .)
$S_{ii} > \sum_{j \neq i} |S_{ij}|$
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 $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, ..., \mathbf{z}_N$ 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 take an arbitrary nonzero 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. Since $\mathbf{x}$ 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.
Each $\mathbf{z}_k\, \mathbf{z}_k^\top$ is a positive real since each $\mathbf{z}_k$ is a real nonzero vector.
The eigenvalues λ_k are all positive.
And every b_k^2 is of course non-negative.
Furthermore, at least one b_k is nonzero, so at least one 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 $\mathbf{w}$ the real part of $\mathbf{z}$, and we call $\mathbf{r}$ the imaginary part of $\mathbf{z}$.

We define these sets of short vectors.
$\mathcal{H} = \{\mathbf{y} | \, \mathbf{y}\, \Psi^\top \text{ is tidy}\}$ $\quad$ $\mathcal{K} = \{\mathbf{y} | \, \mathbf{y}\, \Psi^\top = 0\}$
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 $\Psi^\top$ 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 $\mathcal{K}$.

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

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

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 $\mathcal{E} = \{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, ..., \boldsymbol{\eta}_\ell\}$ 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 $\boldsymbol{\eta}_j$ 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 $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, ..., \boldsymbol{\eta}_\ell\}$ be a real orthonormal basis of $\mathcal{K}$. The induction step gives us a new basis vector $\boldsymbol{\eta}_{\ell+1}$, and adding that vector to the basis gives us a basis of the whole of $\mathcal{H}$. We see that $\boldsymbol{\eta}_{\ell+1}\,\Psi^\top$ is a nonzero tidy real vector, so $\boldsymbol{\eta}_{\ell+1}\,\Psi^\top = \beta\,\mathbf{e}$, where β is a nonzero real scalar. If β is negative, let's change $\boldsymbol{\eta}_{\ell+1}$, 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 $\boldsymbol{\eta}$ to mean the vector in our basis of $\mathcal{H}$ that is not in $\mathcal{K}$.
If $\mathcal{H} \neq \mathcal{K}$ then
$\boldsymbol{\eta}\,\Psi^\top = \beta\,\mathbf{e}$.
$\beta > 0$
If $\mathcal{H} = \mathcal{K}$ then there is no vector $\boldsymbol{\eta}$ and no scalar β .

Let $\mathcal{E} = \{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, ..., \boldsymbol{\eta}_\ell\}$ be our orthonormal basis of $\mathcal{H}$.
We define the real symmetric matrix H .

$$H = \sum_{j=1}^{\ell} \boldsymbol{\eta}_j^\top \boldsymbol{\eta}_j \quad . \tag{8}$$

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

Note that $\boldsymbol{\eta}_j H = \boldsymbol{\eta}_j$ for any basis vector $\boldsymbol{\eta}_j$.
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
$K = \sum_j \boldsymbol{\eta}_j^\top \boldsymbol{\eta}_j$,
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 $\mathbf{y}\,K \;=\; \mathbf{y}$. We also have these.

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

If $\mathcal{H} \neq \mathcal{K}$ then $H \;=\; K + \boldsymbol{\eta}^\top \boldsymbol{\eta}$,
where $\boldsymbol{\eta}$ 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 $\mathbf{y}$ is a happy vector then $\mathbf{y}\,H \;=\; \mathbf{y}$ and $H\,\mathbf{y}^\top =\; \mathbf{y}^\top$.
> If $\mathbf{y}$ 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 $\boldsymbol{\eta}_k$ 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 $\mathbf{x}$ is a short vector. We have $\mathbf{x}\,H\,\Psi^\top =\; \mathbf{x}\,(\sum_k \boldsymbol{\eta}_k^\top \boldsymbol{\eta}_k)\,\Psi^\top =\; \sum_k (\mathbf{x}\,\boldsymbol{\eta}_k^\top)(\boldsymbol{\eta}_k\,\Psi^\top)$.
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 susbsection, $\mathbf{x}$ 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 $\mathbf{x}$ 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 $\mathcal{H} = \mathcal{K}$ then $B_\delta = A_\delta$, so B_δ is positive definite.
If $\mathcal{H} \neq \mathcal{K}$ then $B_\delta = A_\delta + \boldsymbol{\eta}^\top \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is the happy basis vector that isn't in $\mathcal{K}$.
$\mathbf{x}\, B_\delta\, \mathbf{x}^\top = \mathbf{x}\, A_\delta\, \mathbf{x}^\top + (\mathbf{x}\,\boldsymbol{\eta}^\top)(\boldsymbol{\eta}\,\mathbf{x}^\top)$
The second term on the right is non-negative, and A_δ is positive definite, so again B_δ is positive definite.
We conclude: If $\delta > 0$ then B_δ is positive definite.

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.

3.6 The Undiscounted Case

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 parameter.

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 $\boldsymbol{\eta}_k$ be a basis vector of $\mathcal{H}$. Then $\boldsymbol{\eta}_k \Psi^\top$ is a tidy vector, so
$\boldsymbol{\eta}_k \Psi^\top Y = 0$ and $\boldsymbol{\eta}_k \Psi^\top D\,\mathbf{a}^\top = 0$.
Therefore, $H\,\Psi^\top Y = 0$ and $H\,\Psi^\top D\,\mathbf{a}^\top = 0$.

We look at the sequence of successive column weights vectors.
$(\mathbf{v}^{(0)})^\top,\ (\mathbf{v}^{(1)})^\top,\ (\mathbf{v}^{(2)})^\top,\ (\mathbf{v}^{(3)})^\top,\ \ldots\ldots.$
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.
$(\mathbf{x}^{(n+1)})^\top + \mathbf{y}^\top = (Q + \varepsilon\,H)\,(\,(\mathbf{x}^{(n)})^\top + \mathbf{y}^\top) + \varepsilon\,\Psi^\top D\,\mathbf{a}^\top$
We use $Q\,\mathbf{y}^\top = (1-\varepsilon)\,\mathbf{y}^\top$ and $H\,\mathbf{y}^\top = \mathbf{y}^\top$ and $H\,(\mathbf{x}^{(n)})^\top = 0$. We obtain
$(\mathbf{x}^{(n+1)})^\top = Q\,(\mathbf{x}^{(n)})^\top + \varepsilon\,\Psi^\top D\,\mathbf{a}^\top$.
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
$(\mathbf{x}^{(n)})^\top = \varepsilon\,(I - Q^n)\,(I - Q)^{-1}\,\Psi^\top D\,\mathbf{a}^\top + Q^n(\mathbf{x}^{(0)})^\top$.
Since the spectral radius of Q is less than 1, we have $\lim_{n\to\infty} Q^n = 0$ and
$\lim_{n\to\infty}(\mathbf{x}^{(n)})^\top = \varepsilon\,(I - Q)^{-1}\,\Psi^\top D\,\mathbf{a}^\top = B_0^{-1}\,\Psi^\top D\,\mathbf{a}^\top$.
We define

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

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 $\mathbf{u}^\top + \mathbf{y}^\top$.

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, $\mathbf{y}\,\mathbf{u}^\top = \mathbf{y}\,B_0^{-1}\,\Psi^\top D\,\mathbf{a}^\top = \mathbf{y}\,\Psi^\top D\,\mathbf{a}^\top = \alpha\,\mathbf{e}\,D\mathbf{a}^\top = 0$.

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

We look at the sequence of successive column weights vectors. $(\mathbf{v}^{(0)})^\top, (\mathbf{v}^{(1)})^\top, (\mathbf{v}^{(2)})^\top, \ldots.$. 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 $\mathbf{u}_\delta^\top = \ \varepsilon\,(I - Q_\delta)^{-1}\Psi^\top D\,\mathbf{a}^\top$.

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

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

If we begin with $(\mathbf{v}^{(0)})^\top = \ \mathbf{x}^\top + \mathbf{y}^\top$, where $\mathbf{x}^\top$ is the antikernel part and $\mathbf{y}^\top$ is the kernel part, then the limit is $\mathbf{u}_\delta^\top + \mathbf{y}^\top$. We can show directly that $\mathbf{u}_\delta^\top$ is an antikernel vector, for suppose $\mathbf{y}$ is a kernel vector. Then we have $\mathbf{y}A_\delta = \mathbf{y}$, $\mathbf{y}A_\delta^{-1} = \mathbf{y}$, and $\mathbf{y}\,\mathbf{u}_\delta^\top = \ 0$.

4 Decreasing δ

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.[18] 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 $\mathbf{u}_\delta^\top$?

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 λ of A_δ

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

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 $\boldsymbol{\eta}$ 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 $(\hat{N}-1) \times (\hat{N}-1)$ block. The 1×1 block is just the number 1. The

[18] If matrix A is nonsingular then $A^{-1} \ = \ |A|^{-1}\,\mathrm{adj}(A)$. By $\mathrm{adj}(A)$ I mean the adjugate of A .

matrix A_0 is similarly block diagonal with the same $(\hat{N}-1)\times(\hat{N}-1)$ 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 tiniest 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 $\hat{N}-1$. So M_δ has a nonzero $(\hat{N}-1)\times(\hat{N}-1)$ minor, and a nonzero cofactor. So $\operatorname{adj}(M_\delta)$ is not the zero matrix.[19]

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.)

Let the vector $\mathbf{z}$ be the normalized i'th row of $\operatorname{adj}(M_\delta)$ and let $\ddot{\mathbf{z}}^\top$ be the normalized j'th column. So $\mathbf{z}$ and $\ddot{\mathbf{z}}^\top$ 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 $M_\delta\,(\operatorname{adj}(M_\delta)) = |M_\delta|\,I = 0$, we see that $\ddot{\mathbf{z}}^\top$ 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 $\boldsymbol{\eta} = \alpha\,\mathbf{z}_0$. Since $\mathbf{z}_0$ and $\boldsymbol{\eta}$ 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 $\boldsymbol{\zeta}$ 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 $\boldsymbol{\zeta} = \boldsymbol{\eta}$.

In exactly the same way, we define the column vector $\ddot{\boldsymbol{\zeta}}^\top$. 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 $\ddot{\boldsymbol{\zeta}}^\top = \boldsymbol{\eta}^\top$.

So $\boldsymbol{\zeta}\,\ddot{\boldsymbol{\zeta}}^\top$ 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 $\boldsymbol{\zeta}\,\ddot{\boldsymbol{\zeta}}^\top > 0$. We define $\hat{\alpha} = (\boldsymbol{\zeta}\,\ddot{\boldsymbol{\zeta}}^\top)^{-1}$.
So $\hat{\alpha}$ is positive, and if $\delta = 0$ then $\hat{\alpha} = 1$.

[19] By $\operatorname{adj}(M_\delta)$ I mean the adjugate of M_δ .

4.3 Limit of Projected Eigenvector

We define a projection G onto the subspace of vectors orthogonal* to $\boldsymbol{\eta}$.
$G \;=\; I - \boldsymbol{\eta}^\top\boldsymbol{\eta}$.
We are interested in the projection of the vector $\frac{1}{\lambda}\boldsymbol{\zeta}$ onto that subspace. What happens to the projected vector $\frac{1}{\lambda}\boldsymbol{\zeta}\,G$ as $\delta \to 0$? As δ decreases, the vector $\frac{1}{\lambda}\boldsymbol{\zeta}$ gets longer and longer, but it also gets more and more orthogonal to the subspace. So, does the projected vector get longer and longer, or does it get shorter and shorter? In this subsection, we show that it converges to a finite vector, but one that is usually not zero.

We define the vector
$\tau \;=\; \tilde{\mathbf{p}}\Psi$.
The real scalar $\boldsymbol{\zeta}\,\boldsymbol{\eta}^\top$ is a continuous function of δ , and it is 1 when $\delta = 0$. So if we reduce ν yet again we can ensure that $\boldsymbol{\zeta}\,\boldsymbol{\eta}^\top$ is positive for all δ in the interval $[0, \nu]$. We define the vector
$\mathbf{v} \;=\; (\boldsymbol{\zeta}\,\boldsymbol{\eta}^\top)^{-1}\boldsymbol{\zeta}$.
$\mathbf{v}\,\boldsymbol{\eta}^\top \;=\; 1$
And $\mathbf{v}$ 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 $\boldsymbol{\eta}\,\Psi^\top \;=\; \beta\,\mathbf{e}$, we have
$\boldsymbol{\eta}\,A_\delta \;=\; \delta\,\boldsymbol{\eta}\,\Psi^\top F\,\Psi \;=\; \delta\,\beta\,\mathbf{e}\,F\Psi \;=\; \delta\,\beta\,\tilde{\mathbf{p}}\Psi \;=\; \delta\,\beta\,\boldsymbol{\tau}$ and
$A_\delta\,\boldsymbol{\eta}^\top \;=\; \delta\,\Psi^\top F\,\Psi\,\boldsymbol{\eta}^\top \;=\; \delta\,\beta\,\Psi^\top F\,\mathbf{e}^\top \;=\; \delta\,\beta\,\Psi^\top\tilde{\mathbf{p}}^\top \;=\; \delta\,\beta\,\boldsymbol{\tau}^\top$.

$$\boldsymbol{\eta}\,A_\delta \;=\; \delta\,\beta\,\boldsymbol{\tau} \qquad\text{and}\qquad A_\delta\,\boldsymbol{\eta}^\top \;=\; \delta\,\beta\,\boldsymbol{\tau}^\top \tag{9}$$

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

$$\lambda \;=\; \delta\,\beta\,(\mathbf{v}\,\boldsymbol{\tau}^\top) \tag{10}$$

We have $\boldsymbol{\eta}\,\Psi^\top \;=\; \beta\,\mathbf{e}$. If we multiply by $\tilde{\mathbf{p}}^\top$ on the right, we obtain
$\boldsymbol{\eta}\,\boldsymbol{\tau}^\top \;=\; \beta$,
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, (10) 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
$\boldsymbol{\omega}_\delta \;=\; \mathbf{v} - (\mathbf{v}\,\boldsymbol{\tau}^\top)^{-1}\boldsymbol{\tau}$.
This vector is a continuous function of δ , since $\mathbf{v}$ is.
From (9) and the definition of G , we have
$G\,A_\delta \;=\; A_\delta - \boldsymbol{\eta}^\top(\boldsymbol{\eta}\,A_\delta) \;=\; A_\delta - \delta\,\beta\,\boldsymbol{\eta}^\top\boldsymbol{\tau}$
$\mathbf{v}\,G\,A_\delta \;=\; \mathbf{v}\,A_\delta - \delta\,\beta\,(\mathbf{v}\,\boldsymbol{\eta}^\top)\,\boldsymbol{\tau} \;=\; \lambda\,\mathbf{v} - \delta\,\beta\,\boldsymbol{\tau}$
Using (10) and the last equation gives us
$\lambda\,\boldsymbol{\omega}_\delta \;=\; \lambda\,\mathbf{v} - \lambda\,(\mathbf{v}\,\boldsymbol{\tau}^\top)^{-1}\boldsymbol{\tau} \;=\; \lambda\,\mathbf{v} - \delta\,\beta\,\boldsymbol{\tau} \;=\; \mathbf{v}\,G\,A_\delta$.
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 $\delta \to 0$.
Snce B_δ is nonsingular whether δ is zero or not, we have the limit $B_\delta^{-1} \to B_0^{-1}$. We also have
$\boldsymbol{\omega}_0 \;=\; \boldsymbol{\eta} - \beta^{-1}\boldsymbol{\tau}$. Therefore,

$$\lim_{\delta\to 0} \tfrac{1}{\lambda}\,\boldsymbol{\zeta}\,G \;=\; \boldsymbol{\omega}_0\,B_0^{-1} \quad . \tag{11}$$

4.4 Tidying up

Still assuming $\delta > 0$, we can show
$(A_\delta + (1-\lambda)\,\hat{\alpha}\,\ddot{\boldsymbol{\zeta}}^\top\boldsymbol{\zeta})\,(A_\delta^{-1} + (1-\frac{1}{\lambda})\,\hat{\alpha}\,\ddot{\boldsymbol{\zeta}}^\top\boldsymbol{\zeta}) \;=\; I$

simply by multiplying out and using these facts.

$A_\delta \, A_\delta^{-1} \;=\; I \qquad A_\delta \ddot{\boldsymbol{\zeta}}^\top = \; \lambda \, \ddot{\boldsymbol{\zeta}}^\top \qquad \boldsymbol{\zeta} \, A_\delta^{-1} \;=\; \frac{1}{\lambda} \boldsymbol{\zeta} \qquad \boldsymbol{\zeta} \, \ddot{\boldsymbol{\zeta}}^\top = \; \hat{\alpha}$

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.

$A_\delta^{-1} G \;=\; (A_\delta + (1 - \lambda) \, \hat{\alpha} \, \ddot{\boldsymbol{\zeta}}^\top \boldsymbol{\zeta})^{-1} G - \hat{\alpha} \, \ddot{\boldsymbol{\zeta}}^\top \boldsymbol{\zeta} \, G + \hat{\alpha} \, \ddot{\boldsymbol{\zeta}}^\top (\frac{1}{\lambda} \, \boldsymbol{\zeta} \, G)$

We now let $\delta \to 0$. We have the following limits.

$\lambda \to 0 \qquad \hat{\alpha} \to 1 \qquad \ddot{\boldsymbol{\zeta}}^\top \to \boldsymbol{\eta}^\top \qquad \boldsymbol{\zeta} \to \boldsymbol{\eta} \qquad A_\delta \to A_0$

$(A_\delta + (1 - \lambda) \, \hat{\alpha} \, \ddot{\boldsymbol{\zeta}}^\top \boldsymbol{\zeta}) \to B_0$ and B_0 is nonsingular, so $(A_\delta + (1 - \lambda) \, \hat{\alpha} \, \ddot{\boldsymbol{\zeta}}^\top \boldsymbol{\zeta})^{-1} \to B_0^{-1}$.

From these limits and (11) we see that our nice equation becomes

$\lim_{\delta \to 0} A_\delta^{-1} G \;=\; B_0^{-1} G + \boldsymbol{\eta}^\top \boldsymbol{\omega}_0 \, B_0^{-1}$.

We have $\boldsymbol{\eta} \, B_0 \;=\; \boldsymbol{\eta}$, and so $\boldsymbol{\eta} \;=\; \boldsymbol{\eta} \, B_0^{-1}$.

$\boldsymbol{\omega}_0 \, B_0^{-1} \;=\; \boldsymbol{\eta} - \beta^{-1} \boldsymbol{\tau} \, B_0^{-1}$

$\lim_{\delta \to 0} A_\delta^{-1} G \;=\; B_0^{-1} G + \boldsymbol{\eta}^\top \boldsymbol{\eta} - \beta^{-1} \boldsymbol{\eta}^\top \boldsymbol{\tau} \, B_0^{-1}$

We multiply on the right by $\Psi^\top D \, \mathbf{a}^\top$ and use

$\boldsymbol{\eta} \, \Psi^\top D \, \mathbf{a}^\top = \; \beta \, \mathbf{e} \, D \, \mathbf{a}^\top = \; 0$ and $G \, \Psi^\top D \, \mathbf{a}^\top = \; \Psi^\top D \, \mathbf{a}^\top$.

$\lim_{\delta \to 0} A_\delta^{-1} \Psi^\top D \, \mathbf{a}^\top \;=\; B_0^{-1} \Psi^\top D \, \mathbf{a}^\top - \beta^{-1} \boldsymbol{\eta}^\top \boldsymbol{\tau} \, B_0^{-1} \Psi^\top D \, \mathbf{a}^\top$

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 have

$\lim_{\delta \to 0} \mathbf{u}_\delta^\top \;=\; \mathbf{u}^\top - \beta^{-1} \boldsymbol{\eta}^\top (\boldsymbol{\tau} \, \mathbf{u}^\top)$.

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 The Limit Vector gives us the Value Estimates.

We have seen that the sequence of average steps $(\mathbf{v}^{(0)})^\top$, $(\mathbf{v}^{(1)})^\top$, $(\mathbf{v}^{(2)})^\top$, $(\mathbf{v}^{(3)})^\top$, ... converges to a limit. What the limit is depends in general on the starting vector. I shall call $\mathbf{v}^\top$ a limit vector if it is the limit of some such sequence. It's easy to see that a vector $\mathbf{v}^\top$ is a limit vector if and only if the average change formula (4) has the value zero.

Each weights vector $(\mathbf{v}^{(n)})^\top$ gives us a corresponding vector $\Psi \, (\mathbf{v}^{(n)})^\top$ of estimates of state values. The sequence of steps is supposed to improve the estimates, so in this subsection we shall be interested in estimates vectors $\Psi \, \mathbf{v}^\top$ where $\mathbf{v}^\top$ is a limit vector. In this section 5, when I say estimates I shall mean estimates given by a limit vector.

If $\mathbf{v}$ is *any* short vector, we define $\bar{\mathbf{v}}^\top = \; \Psi \, \mathbf{v}^\top$. So if $\mathbf{v}^\top$ is a limit vector then of course $\bar{\mathbf{v}}^\top$ is the corresponding estimates vector.

We can see that the average state value $\tilde{\mathbf{p}} \, \mathbf{c}_\delta^\top$ is zero, but is the average estimate zero? In some cases we can show that it is. Consider the discounted case. If $\mathbf{v}^\top$ is a limit vector then $\mathbf{v}^\top = \; \mathbf{u}_\delta^\top + \mathbf{y}^\top$ for some kernel vector $\mathbf{y}^\top$. (See subsection 3.7.) Multiplying that equation by Ψ on the left gives us $\bar{\mathbf{v}}^\top = \; \bar{\mathbf{u}}_\delta^\top$. The average estimate is $\tilde{\mathbf{p}} \, \bar{\mathbf{v}}^\top = \; \tilde{\mathbf{p}} \, \bar{\mathbf{u}}_\delta^\top$. The next paragraph shows this is zero provided $\mathcal{H} \neq \mathcal{K}$.

We examine $\tilde{\mathbf{p}} \, \bar{\mathbf{u}}_\delta^\top$, assuming that $\mathcal{H} \neq \mathcal{K}$.

Remember that the definitions of β and $\boldsymbol{\tau}$ give us these three facts.

$\boldsymbol{\eta} \, \Psi^\top = \; \beta \, \mathbf{e} \qquad \beta > 0 \qquad \boldsymbol{\tau} \;=\; \tilde{\mathbf{p}} \, \Psi$ First let's assume $\delta > 0$.

We have $\mathbf{e} \, Y_\delta \;=\; \tilde{\mathbf{p}} \, (I - (1 - \delta) \, P) \;=\; \delta \, \tilde{\mathbf{p}}$, and so

$\boldsymbol{\eta} \, \Psi^\top Y_\delta \, \Psi \;=\; \beta \, \mathbf{e} Y_\delta \Psi \;=\; \beta \, \delta \, \tilde{\mathbf{p}} \, \Psi \;=\; \beta \, \delta \, \boldsymbol{\tau}$.

Since $\boldsymbol{\eta}$ is an antikernel vector, $\boldsymbol{\eta} \, K \;=\; 0$, so

$\boldsymbol{\eta} \, A_\delta \;=\; \beta \, \delta \, \boldsymbol{\tau}$. $\qquad \beta \, \delta \, \boldsymbol{\tau} \, A_\delta^{-1} \;=\; \boldsymbol{\eta}$

$\beta \, \delta \, \boldsymbol{\tau} \, A_\delta^{-1} \, \Psi^\top D \, \mathbf{a}^\top = \; \boldsymbol{\eta} \, \Psi^\top D \mathbf{a}^\top = \; \beta \, \mathbf{e} \, D \, \mathbf{a}^\top = \; \beta \, \tilde{\mathbf{p}} \, \mathbf{a}^\top = \; 0$

$\beta \, \delta \, \boldsymbol{\tau} \, \mathbf{u}_\delta^\top = \; 0$ Since β and δ are both positive, we have

$\boldsymbol{\tau} \, \mathbf{u}_\delta^\top = \; 0$. This gives us

$\tilde{\mathbf{p}} \, \bar{\mathbf{u}}_\delta^\top = \; \tilde{\mathbf{p}} \, \Psi \, \mathbf{u}_\delta^\top = \; \boldsymbol{\tau} \, \mathbf{u}_\delta^\top = \; 0$.

That takes care of the $\delta > 0$ case. For the $\delta = 0$ case we use

$\boldsymbol{\tau} \, \boldsymbol{\eta}^\top = \; \tilde{\mathbf{p}} \, \Psi \, \boldsymbol{\eta}^\top = \; \tilde{\mathbf{p}} \, (\beta \, \mathbf{e}^\top) \;=\; \beta$.

$\tilde{\mathbf{p}}\,\bar{\mathbf{u}}_0^\top = \boldsymbol{\tau}\,\mathbf{u}_0^\top = \boldsymbol{\tau}\,(\mathbf{u}^\top - (\boldsymbol{\tau}\,\mathbf{u}^\top)\,\beta^{-1}\boldsymbol{\eta}^\top) = \boldsymbol{\tau}\,\mathbf{u}^\top - (\boldsymbol{\tau}\,\mathbf{u}^\top)\,\beta^{-1}(\boldsymbol{\tau}\,\boldsymbol{\eta}^\top) = 0$
So we see that if $\mathcal{H} \neq \mathcal{K}$ then in all cases we have $\tilde{\mathbf{p}}\,\bar{\mathbf{u}}_\delta^\top = 0$.

The situation in the undiscounted case is rather different. If $\mathbf{v}^\top$ is a limit vector then $\mathbf{v}^\top = \mathbf{u}^\top + \mathbf{y}^\top$ for some happy $\mathbf{y}^\top$. (See subsection 3.6.) Multiplying that equation by Ψ on the left gives us $\bar{\mathbf{v}}^\top = \bar{\mathbf{u}}^\top + \chi\,\mathbf{e}^\top$ for some real parameter χ . The average estimate is $\tilde{\mathbf{p}}\,\bar{\mathbf{v}}^\top = \tilde{\mathbf{p}}\,\bar{\mathbf{u}}^\top + \chi$. If $\mathcal{H} \neq \mathcal{K}$ then the parameter depends on the starting vector and might be anything. The average estimate is unlikely to be zero.

In the undiscounted case, the limit vector is $\mathbf{u}^\top + \mathbf{y}^\top$. So the estimated value of state i is $\bar{u}_i + \chi$. The $\bar{u}_i$ is given by the sad vector $\mathbf{u}^\top$ and the χ is given by the happy vector $\mathbf{y}^\top$. ($\chi = \mathbf{e}_i \Psi\,\mathbf{y}^\top$) The χ is the same for every state.
In the discounted case with $\delta \to 0$, the limit vector is $\mathbf{u}_0^\top$ plus a kernel vector. This is $\mathbf{u}^\top - (\boldsymbol{\tau}\,\mathbf{u}^\top)\,\beta^{-1}\boldsymbol{\eta}^\top$ plus a kernel vector if $\mathcal{H} \neq \mathcal{K}$. It's like the undiscounted case except that now the happy vector is $-(\boldsymbol{\tau}\,\mathbf{u}^\top)\,\beta^{-1}\boldsymbol{\eta}^\top$ plus a kernel vector. So the estimated value of of state i is $\bar{u}_i - \tilde{\mathbf{p}}\,\bar{\mathbf{u}}^\top$, since $\mathbf{e}_i\Psi\,((\boldsymbol{\tau}\,\mathbf{u}^\top)\,\beta^{-1}\boldsymbol{\eta}^\top) = (\boldsymbol{\tau}\,\mathbf{u}^\top)\,\beta^{-1}(\mathbf{e}_i\Psi\,\boldsymbol{\eta}^\top) = (\tilde{\mathbf{p}}\,\bar{\mathbf{u}}^\top)\,\beta^{-1}(\mathbf{e}_i(\beta\,\mathbf{e}^\top)) = \tilde{\mathbf{p}}\,\bar{\mathbf{u}}^\top$. The happy vector is just what it takes to make the average estimate zero.

5.2 Adjusting the Estimates – False Values

Given any estimates vector $\bar{\mathbf{v}}^\top$, we can *adjust* the estimates by subtracting the average estimate from each estimate. We use the following notation. If $\mathbf{v}$ is any short vector we define $\ddot{\mathbf{v}}^\top = (I - \mathbf{e}^\top\tilde{\mathbf{p}})\,\Psi\,\mathbf{v}^\top$. This means that for any short vector $\mathbf{v}$ we have $\ddot{\mathbf{v}}^\top = (I - \mathbf{e}^\top\tilde{\mathbf{p}})\,\bar{\mathbf{v}}^\top$.
Suppose we *adjust* the vector $\bar{\mathbf{v}}^\top$ by subtracting the average of the entries from every entry. Then we obtain the corresponding adjusted vector $\bar{\mathbf{v}}^\top - (\tilde{\mathbf{p}}\,\bar{\mathbf{v}}^\top)\,\mathbf{e}^\top$, which is $\ddot{\mathbf{v}}^\top$. Of course we always have $\tilde{\mathbf{p}}\,\ddot{\mathbf{v}}^\top = 0$. If $\mathbf{v}^\top$ is a limit vector then $\bar{\mathbf{v}}^\top$ is the corresponding estimates vector and $\ddot{\mathbf{v}}^\top$ is the corresponding adjusted estimates vector. In some cases the adjustment is unnecessary. For example, note that $(a_i - \ddot{v}_i + \ddot{v}_j) = (a_i - \bar{v}_i + \bar{v}_j)$, so it makes no difference to undiscounted linear-TD(0) whether it uses $\ddot{\mathbf{v}}^\top$ or $\bar{\mathbf{v}}^\top$ when it adjusts the weights.

We saw that in the undiscounted case, if $\mathbf{v}^\top$ is a limit vector then we have $\bar{\mathbf{v}}^\top = \bar{\mathbf{u}}^\top + \chi\,\mathbf{e}^\top$. Multiplying by $(I - \mathbf{e}^\top\tilde{\mathbf{p}})$ on the left gives us $\ddot{\mathbf{v}}^\top = \ddot{\mathbf{u}}^\top$. So the vector of adjusted estimates is $\ddot{\mathbf{u}}^\top$. The vector of adjusted estimates is independent of the starting vector.

We saw that in the discounted case, if $\mathbf{v}^\top$ is a limit vector then we have $\bar{\mathbf{v}}^\top = \bar{\mathbf{u}}_\delta^\top$. Multiplying by $(I - \mathbf{e}^\top\tilde{\mathbf{p}})$ on the left gives us $\ddot{\mathbf{v}}^\top = \ddot{\mathbf{u}}_\delta^\top$. So the vector of adjusted estimates is $\ddot{\mathbf{u}}_\delta^\top$. Here too, it's independent of the starting vector.

So we see that although in general the limit vector depends on the starting vector, the adjusted estimates vector is unique and the same for all limit vectors, the same for all starting vectors. If $\mathbf{v}^\top$ is any limit vector, then $\ddot{\mathbf{v}}^\top$ is the unique vector of adjusted value estimates.

I shall call $\ddot{\mathbf{v}}^\top$ the vector of *false values*, since its entries are not the true values. They are only estimates. If $\mathbf{v}^\top$ is any limit vector then $\ddot{\mathbf{v}}^\top$ is the unique false values vector. We have seen that the false values vector is $\ddot{\mathbf{u}}^\top$ in the undiscounted case, and $\ddot{\mathbf{u}}_\delta^\top$ in the discounted case.

We have been implying that in general the false values are good estimates, and in a sense they are, but of course they are rarely equal to the true state values $\mathbf{c}_\delta^\top$.

But suppose we are so lucky as to have $\mathbf{c}_\delta \in \mathrm{Ran}(\Psi^\top)$.
Then there is a lucky vector $\mathbf{v}^\top$ of weights such that $\Psi\,\mathbf{v}^\top = \mathbf{c}_\delta^\top$, and so $\bar{\mathbf{v}}^\top = \mathbf{c}_\delta^\top$.
Then the average change formula (4) is zero since $(I - (1-\delta)P)\,\mathbf{c}_\delta^\top = \mathbf{a}^\top$ and $Y_\delta\mathbf{c}_\delta^\top = D\,\mathbf{a}^\top$.
So the lucky $\mathbf{v}^\top$ is a limit vector, and $\ddot{\mathbf{v}}^\top$ is the false values vector.
Since $\tilde{\mathbf{p}}\,\mathbf{c}_\delta^\top = 0$, we can multiply $\bar{\mathbf{v}}^\top = \mathbf{c}_\delta^\top$ on the left by $(I - \mathbf{e}^\top\tilde{\mathbf{p}})$ and obtain $\ddot{\mathbf{v}}^\top = \mathbf{c}_\delta^\top$. The false values are the same as the true values. So in this case the estimates are correct. This argument works in both the discounted and undiscounted case. If there exist one or more weights vectors that give completely correct estimates (if $\mathbf{c}_\delta \in \mathrm{Ran}(\Psi^\top)$) then linear-TD(0) will find one of them.

If $\Psi = I$ then linear-TD(0) becomes what the Evolutionary Computation literature calls a *simple bucket brigade.*[20] Then for any weights vector $\mathbf{v}^\top$ we have $\mathbf{v}^\top = \bar{\mathbf{v}}^\top$, so the weights are the unadjusted state value estimates. The weights are called cash balances. The simple bucket brigade is undiscounted, but in this paragraph we allow it to be either undiscounted or discounted. Let $\mathbf{v}^\top$ be any limit vector. Since $\mathrm{Ran}(\Psi^\top)$ is the whole space, we have $\mathbf{c}_\delta \in \mathrm{Ran}(\Psi^\top)$, so the false values $\ddot{\mathbf{v}}^\top$ are the same as the true values $\mathbf{c}_\delta^\top$. Since $\mathbf{v}^\top = \bar{\mathbf{v}}^\top$, if we adjust the cash balances $\mathbf{v}^\top$ by subtracting from each the average

[20] Sometimes the simple bucket brigade is called “the bucket brigade on a Markov chain”.

cash balance, we obtain $\ddot{\mathbf{v}}^\top$, which is $\mathbf{c}_\delta^\top$. The simple bucket brigade converges to the correct values.[21]

5.3 False Payoffs in the Undiscounted Case

In this subsection we discuss only the undiscounted case. In the undiscounted case, the false values vector is $\ddot{\mathbf{u}}^\top$. Each false value $\ddot{u}_i$ is of course an estimate of the true value c_i . The estimates can be very biased and wrong. To call them approximations would be misleading.[22]

We have $\tilde{\mathbf{p}}\,\ddot{\mathbf{u}}^\top = 0$, and we define

$\hat{\mathbf{a}}^\top = (I - P)\,\ddot{\mathbf{u}}^\top$.

I call $\hat{a}_i$ the *false payoff* of i .

Since $(I - P)\,\ddot{\mathbf{u}}^\top = (I - P)\,(I - \mathbf{e}^\top\tilde{\mathbf{p}})\,\bar{\mathbf{u}}^\top = (I - P)\,\bar{\mathbf{u}}^\top$, we have

$\hat{\mathbf{a}}^\top = (I - P)\,\Psi\mathbf{u}^\top$.

We define the matrix

$\hat{P} = I - P + \mathbf{e}^\top\tilde{\mathbf{p}}$.

We note that $\hat{P}$ is nonsingular.[23] 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

$\mathbf{c}^\top = \hat{P}^{-1}\,\mathbf{a}^\top = \sum\!\!\!\mathrm{C}_{n=0}^{\infty}(P^n\,\mathbf{a}^\top)$.

We can think of $\hat{P}^{-1}\,\mathbf{a}^\top$ as an alternative definition of $\mathbf{c}^\top$.

We also have $\hat{P}\,\ddot{\mathbf{u}}^\top = \hat{\mathbf{a}}^\top$ and $\tilde{\mathbf{p}}\,\hat{\mathbf{a}}^\top = 0$, so $\hat{P}\,\sum\!\!\!\mathrm{C}_{n=0}^{\infty}(P^n\,\hat{\mathbf{a}}^\top) = \hat{\mathbf{a}}^\top$ and

$\ddot{\mathbf{u}}^\top = \hat{P}^{-1}\,\hat{\mathbf{a}}^\top = \sum\!\!\!\mathrm{C}_{n=0}^{\infty}(P^n\,\hat{\mathbf{a}}^\top)$.

(The Cesaro sum converges for the same reasons $\sum\!\!\!\mathrm{C}_{n=0}^{\infty}(P^n\,\mathbf{a}^\top)$ coverges.[24])

The false values $\ddot{\mathbf{u}}^\top$ 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 $\ddot{\mathbf{u}}^\top$ given to us by linear-TD(0) rather than on the true values $\mathbf{c}^\top$, it's as if linear-TD(0) has changed the true payoffs $\mathbf{a}^\top$ into false payoffs $\hat{\mathbf{a}}^\top$. This change is a problem inherent in all temporal difference methods.[25] What makes this all non-trivial is that there is an interesting relation between the true payoffs $\mathbf{a}^\top$ and the false payoffs $\hat{\mathbf{a}}^\top$, which we now show.

Since $\mathbf{u}$ is sad, we have , $H\,\mathbf{u}^\top = 0$, so

$\Psi^\top D\,\mathbf{a}^\top = (B_0)\,(B_0^{-1}\,\Psi^\top D\,\mathbf{a}^\top) = (\Psi^\top Y\,\Psi + H)\,(\mathbf{u}^\top) = \Psi^\top Y\,\Psi\,\mathbf{u}^\top$.

$$\Psi^\top D\,\mathbf{a}^\top = \Psi^\top Y\,\Psi\,\mathbf{u}^\top \tag{12}$$

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

$$\Psi^\top D\,\hat{\mathbf{a}}^\top = \Psi^\top D\,\mathbf{a}^\top \qquad\qquad \sum_i \tilde{p}_i\psi_{ik}\hat{a}_i = \sum_i \tilde{p}_i\psi_{ik}a_i \tag{13}$$

In subsection 2.2 we said that if i is a state then ψ_{ik} is said to be the value of its k'th feature. So we can think of $\sum_i \tilde{p}_i\psi_{ik}a_i$ as the average payoff allocated to feature k . We see from (13) 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 (13). 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

$\mathbf{z}^\top = (I - P)\,\Psi\,\mathbf{x}^\top$.

[21] That statement was proved directly in [11] for the undiscounted case.

[22] There is a simple discussion of the biases in [10]. Simple because the discussion is only of bucket brigades that are special cases of linear-TD(0).

[23] Suppose $\mathbf{v}\hat{P} = 0$ 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 Perron-Frobenius theorem [2] tells us that the space of such eigenvectors is one dimensional. Therefore, $\mathbf{v} = \lambda\tilde{\mathbf{p}}$ for some scalar λ . Therefore, $\mathbf{v} = \lambda\tilde{\mathbf{p}} = \lambda\tilde{\mathbf{p}}\hat{P} = \mathbf{v}\hat{P} = 0$. Contradiction.

[24] Change the payoffs to $\hat{\mathbf{a}}^\top$ and then use the proof in [13].

[25] Or if you prefer, the change is one way of viewing a problem that is inherent in all temporal difference methods.

$\Psi^\top D\,\mathbf{z}^\top = \Psi^\top D\,(I-P)\,\Psi\,\mathbf{x}^\top = \Psi^\top Y\,\Psi\,\mathbf{x}^\top = B_0\,\mathbf{x}^\top$

By (12) we have

$\Psi^\top D\,\mathbf{a}^\top = B_0\,\mathbf{u}^\top$.

Since $\Psi^\top D\,\mathbf{z}^\top = \Psi^\top D\,\mathbf{a}^\top$, the previous two equations give us $B_0\,\mathbf{x}^\top = B_0\,\mathbf{u}^\top$, so $\mathbf{x}^\top = \mathbf{u}^\top$.

We then have

$\mathbf{z}^\top = (I-P)\,\Psi\,\mathbf{x}^\top = (I-P)\,\Psi\,\mathbf{u}^\top = \hat{\mathbf{a}}^\top$.

6 Other Linear-TD(0) Methods

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 (5) 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 $\mathbf{y}$ 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 (9) 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 becomes $\delta^{-1}\,\bar{m}\,\mathbf{e}^\top$. Since this is tidy, adjustment will get rid of it. But if δ is small the tidy term is huge, and this can cause calculation problems. We now look at the even worse case when $\delta = 0$.

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. Presumably the sequence doesn't converge unless $\bar{m} = 0$.

6.2 TD(0)-future Methods

We define $\mathbf{b}^\top = P\,\mathbf{a}^\top$. Then we have

$\sum_j F_{ij} b_i = \tilde{p}_i b_i = \tilde{p}_i \sum_j P_{ij} a_j = \sum_j F_{ij} a_j$, and therefore,

$\varepsilon \sum_{ij} F_{ij} b_i \psi_{ik} = \varepsilon \sum_{ij} F_{ij} a_j \psi_{ik}$.

$$\varepsilon \sum_{ij} F_{ij}(b_i - \bar{v}_i + (1-\delta)\,\bar{v}_j)\,\psi_{ik} = \varepsilon \sum_{ij} F_{ij}(a_j - \bar{v}_i + (1-\delta)\,\bar{v}_j)\,\psi_{ik} \quad . \tag{14}$$

There is a class of TD(0) methods that I call TD(0)-future methods.

In these methods, if the state transition is $i \to j$, the increment to v_k is

$\varepsilon\,(a_j - \bar{v}_i + (1-\delta)\,\bar{v}_j)\,\psi_{ik}$

rather than formula (2). Sutton and Barto [8] frequently use these methods.

The TD(0)-future average change in v_k then is the right side of (14) rather than (4). Let's use the left side of (14). That's just (4) 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

$$\mathbf{z}_\delta^\top = A_\delta^{-1}\,\Psi^\top D\,\mathbf{b}^\top \qquad \text{and} \qquad \mathbf{z}^\top = B_0^{-1}\,\Psi^\top D\,\mathbf{b}^\top \ .$$

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 $\mathbf{w}_\delta^\top = \sum_{n=0}^{\infty}(\,(1-\delta)^n P^n\,\mathbf{b}^\top)$. We see that $\mathbf{c}_\delta^\top = \mathbf{a}^\top + (1-\delta)\,\mathbf{w}_\delta^\top$. So we would like our new estimates to be related as follows.

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

We shall see that if $\mathbf{a} \in \mathrm{Ran}(\Psi^\top)$ then we do have that relationship.

Suppose $\mathbf{a} \in \mathrm{Ran}(\Psi^\top)$ and $\delta > 0$.

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

$\Psi(\mathbf{x}^\top + \mathbf{y}^\top) = \mathbf{a}^\top$. And so we have $\Psi\,\mathbf{y}^\top = \mathbf{a}^\top$.

$A_\delta\,\mathbf{y}^\top = (\Psi^\top Y_\delta\,\Psi + K)\,\mathbf{y}^\top = \Psi^\top Y_\delta\,\mathbf{a}^\top = \Psi^\top D\,(I-(1-\delta)\,P)\,\mathbf{a}^\top = \Psi^\top D\,\mathbf{a}^\top - (1-\delta)\,\Psi^\top D\,\mathbf{b}^\top$

We multiply by A_δ^{-1} on the left. $\mathbf{y}^\top = \mathbf{u}_\delta^\top - (1-\delta)\,\mathbf{z}_\delta^\top$

Now we multiply by Ψ on the left. $\mathbf{a}^\top = \bar{\mathbf{u}}_\delta^\top - (1-\delta)\,\bar{\mathbf{z}}_\delta^\top$

So we see that the vector $\mathbf{a}^\top + (1-\delta)\,\ddot{\mathbf{z}}_\delta^\top - \ddot{\mathbf{u}}_\delta^\top$ 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 $\mathbf{a} \in \mathrm{Ran}(\Psi^\top)$ and $\delta > 0$, then $\ddot{\mathbf{u}}_\delta^\top = \mathbf{a}^\top + (1-\delta)\,\ddot{\mathbf{z}}_\delta^\top$.

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

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

$(I-P)\,\Psi\,\mathbf{x}^\top = \mathbf{a}^\top - \mathbf{b}^\top$. We multiply on the left by $\Psi^\top D$.

$\Psi^\top Y\,\Psi\,\mathbf{x}^\top = \Psi^\top D\,(\mathbf{a}^\top - \mathbf{b}^\top)$. Since $H\,\mathbf{x}^\top = 0$, this becomes

$B_0\,\mathbf{x}^\top = \Psi^\top D\,(\mathbf{a}^\top - \mathbf{b}^\top)$.

$\mathbf{x}^\top = B_0^{-1}\,\Psi^\top D\,(\mathbf{a}^\top - \mathbf{b}^\top) = \mathbf{u}^\top - \mathbf{z}^\top$.

$(\mathbf{x}^\top + \mathbf{y}^\top) + \mathbf{z}^\top - \mathbf{u}^\top = \mathbf{y}^\top$ Now multiply by Ψ on the left.

$\mathbf{a}^\top + \bar{\mathbf{z}}^\top - \bar{\mathbf{u}}^\top = \Psi\,\mathbf{y}^\top$

Since $\Psi\,\mathbf{y}^\top$ is a tidy vector, we see that the vector $\mathbf{a}^\top + \ddot{\mathbf{z}}^\top - \ddot{\mathbf{u}}^\top$ 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 $\mathbf{a} \in \mathrm{Ran}(\Psi^\top)$ then $\ddot{\mathbf{u}}^\top = \mathbf{a}^\top + \ddot{\mathbf{z}}^\top$.

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 $\mathbf{e} \in \mathrm{Ran}(\Psi^\top)$, 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 $\mathbf{m} \in \mathrm{Ran}(\Psi^\top)$. If we also have $\mathbf{e} \in \mathrm{Ran}(\Psi^\top)$, then $\mathbf{a} \in \mathrm{Ran}(\Psi^\top)$.

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 bucket brigades that are special cases 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.

We then define what I call false payoffs and use the undiscounted limit formula to prove equation (13), which relates true and false payoffs. This too is new. It gives us a handle on the biases that bedevil linear-TD(0) and other temporal difference methods.

References

[1] S. Bhatnagar, R. S. Sutton, M. Ghavamzadeh, and M. Lee. Natural actor-critic algorithms. *Automatica*, 45(11):2471–2482, 2009.

[2] F. R. Gantmacher. *Applications of the Theory of Matrices.* Interscience Publishers, New York, 1959.

[3] D. O. Hebb. *The Organization of Behavior: A Neuropsychological Theory.* Wiley, New York, 1949.

[4] J. H. Holland, K. J. Holyoak, R. E. Nisbett, and P. R. Thagard. *Induction: Processes of Inference, Learning and Discovery.* MIT Press, Cambridge, MA, 1986.

[5] J. R. Norris. *Markov Chains.* Cambridge University Press, Cambridge, 1997.

[6] A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM J. of Res. Develop.*, 13:210–229, 1959.

[7] R. S. Sutton and A. G. Barto. Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88(2):135–170, 1981.

[8] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction, Second edition.* MIT Press, Cambridge, Mass., 2018.

[9] Richard S. Varga. *Geršgorin and His Circles*, volume SSCM-36. Springer Series in Computational Mathematics, Heidelberg, 2004.

[10] T. H. Westerdale. A defense of the bucket brigade. In J. David Schaffer, editor, *Proceedings of the Third International Conference on Genetic Algorithms*, pages 282–290, San Mateo, CA, 1989. Morgan Kaufmann.

[11] 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.

[12] Tom Westerdale. Bucket Brigade Convergence on Markov Chains. pages 1–50, 2022. unpublished. URL `https://www.dcs.bbk.ac.uk/~tom/convergence.pdf`.

[13] Tom Westerdale. Learning Resembles Evolution – the Markov Case. pages 1–12, 2024. unpublished. URL `https://www.dcs.bbk.ac.uk/~tom/briefsymmetric.pdf`.

[14] Tom Westerdale. Learning Resembles Evolution even when using Temporal Diffference. pages 1–16, 2024. unpublished. URL `https://www.dcs.bbk.ac.uk/~tom/briefgeneral.pdf`.