QUASI-HYPERBOLIC DISCOUNTING NOTE

The aim is to apply ADP methods to quasi-hyperbolic discounting (QHD).

1. Fully Rational QHD

Fully rational means that at time zero you choose a stationary policy that maximizes lifetime value and keep it forever.

Lifetime value of policy σ starting from $t = 0$ is

$$w_\sigma(x) := r_\sigma(x) + \beta \mathbb{E}_x \sum_{t \geqslant 1} \delta^t r_\sigma(X_t)$$

Pointwise this is

$$w_\sigma = r_\sigma + \beta P_\sigma \sum_{t \geqslant 1} (\delta P_\sigma)^t r$$

Suppose for now that states and actions are finite, so w_σ takes values in $V := \mathbb{R}^{\mathsf{X}}$.

We generalize to allow state-dependent discounting, taking B_σ and D_σ to be positive linear operators over V and writing

$$w_\sigma = r_\sigma + B_\sigma \sum_{t \geqslant 1} D_\sigma^t r_\sigma \tag{1}$$

To maximize w_σ using ADP theory we can

(i) write w_σ as the fixed point of an operator S_σ and
(ii) show that $(V, \{S_\sigma\})$ is a globally stable ADP.

Let's start by finding S_σ. For now we drop the subscript σ To simplify notation. Rearranging (1) gives

$$w = r + BDr + B \sum_{t \geqslant 2} D^t r = r + BDr + BD \sum_{t \geqslant 1} D^t r$$

Assuming that B is invertible and using (1) again gives

$$w = r + BDr + BDB^{-1}(w - r)$$

For example, in the case where $B = \beta$, we have

$$w = r + \beta D r + D(w - r) = r - (1 - \beta) D r + D\, w$$

For this case, putting the subscript σ back in, we write

$$S_\sigma\, w = r_\sigma - (1 - \beta) D_\sigma r_\sigma + D_\sigma\, w$$

If, say, $\sup_\sigma \rho(D_\sigma) < 1$, then we have a globally stable ADP with value function v^* and at least one optimal policy σ^*. The usual optimality results apply:

(i) Bellman's principle of optimality holds.
(ii) VFI, HPI, OPI converge, etc.

2. QHD with Limited Self-Control

In HL and BRW, the perspective is as follows:

- There are separate "selves" at each point in time t.
- The $t = 1$ self chooses a policy σ and receives rewards according to

$$v_\sigma = \sum_{t \geqslant 0} (\delta P_\sigma)^t r_\sigma = \sum_{t \geqslant 0} D^t r_\sigma \tag{2}$$

- The $t = 0$ self takes v_σ in (2) as given and chooses a policy τ to solve

$$\tau(x) \in \operatorname*{arg\,max}_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \sum_{x'} v_\sigma(x') P(x, a, x') \right\}$$

We have a stationary Markov Nash equilibrium (SMNE) when $\tau = \sigma$.

We can write this more abstractly as follows: Let

- $T_\sigma\, v = r_\sigma + D_\sigma\, v$
- $\hat{T}_\sigma\, v = r_\sigma + B_\sigma\, v$ and $\hat{T} = \bigvee_\sigma \hat{T}_\sigma$

Let $\tau = M\sigma$ be defined by choosing v_σ as the fixed point of T_σ and then τ such that $\hat{T} v_\sigma = \hat{T}_\tau v_\sigma$. We seek a fixed point of M.

Questions:

(i) The policy τ is not necessarily optimal for the self at $t = 1$. Why would the self at $t = 1$ accept it?

(ii) Why does the self at $t = 1$ have different preferences to the self at $t = 0$? If they are all copies of the same "self," then each faces an infinite horizon and has the same lifetime objective (1). In particular, the self at $t = 1$ should choose σ to maximize (1) rather than (2).

(iii) What justification is there for focusing only on *stationary* Markov Nash equilibria?

(iv) Are there any stability results for SMNE, showing that boundedly rational agents naturally converge to this behavior.

(v) Given that there are no uniqueness results for SMNE, how can we use this for quantitative work?