Theorem (≈6.2.2 Puterman): If there exists a v such that $v = Lv$, then $v = u_\lambda^{\pi^*}$.

This proof is in two parts: (a) prove that if $v \geq Lv$ then $v \geq u_\lambda^{\pi*}$ and (b) prove if $v \leq Lv$ then $v \leq u_\lambda^{\pi*}$. From which it follows that if $v = Lv$ and $v = u_\lambda^{\pi*}$.

Part (a): For some arbitrary policy $\pi = (d, d, d, \ldots.)$ if $v \geq Lv$ then $v \geq \max_d\{r_d + \lambda P_d v\}$. It follows that:

$$\begin{aligned} v &\geq r_d + \lambda P_d v \\ &= r_d + \lambda P_d(r_d + P_d v) \\ &= r_d + \lambda P_d r_d + \cdots + \lambda^n P_d^n v \end{aligned}$$

Therefore it follows from subtracting $u_\lambda^{\pi*}$ from both sides that:

$$v - u_\lambda^{\pi*} \geq \lambda^n P_d^n v - \sum_{k=n}^{\infty} \lambda^k P_d^k r_d \qquad (1)$$

Since $||\lambda^n P_d^n v|| \leq \lambda^n ||v||$ then since $0 < \lambda < 1$ for any ϵ there exists some n sufficiently large:

$$\left(-\frac{\epsilon}{2}\right) e \leq \lambda^n P_d^n v \leq \left(\frac{\epsilon}{2}\right) e$$

Where e is a vector of ones of dimension equal to that of vector v. Now since the rewards are finite and the norm of a transition probability matrix is 1 it follows that:

$$||\lambda^k P_d^k r_d|| \leq \lambda^k ||P r_d|| \leq \lambda^k ||r_d|| \leq \lambda^k M$$

where M is a finite upper bound on all elements of the reward vector. Thus, it follows that

$$\sum_{k=n}^{\infty} \lambda^k P_d^k r_d \leq M \sum_{k}^{\infty} \lambda^k = \frac{\lambda^n}{1-\lambda} M$$

Where the latter equality is obtained by taking the difference between $\sum_{i=0}^{\infty} \lambda^i = \frac{1}{1-\lambda}$ and $\sum_{i=0}^{n} \lambda^i = \frac{1-\lambda^n}{1-\lambda}$.

Using this bound we see that for any given ϵ there exists n sufficiently large such that:

$$\left(-\frac{\epsilon}{2}\right) e \leq \sum_{k=0}^{\infty} \lambda^k P_d^k r_d \leq \left(\frac{\epsilon}{2}\right) e$$

Using equation (1) and the above results we have that $v \geq u_\lambda^{\pi*} + \lambda^n P_d^n v - \sum_{k=n}^{\infty} \lambda^k P_d^k r_d \geq u_\lambda^{\pi*} - \epsilon$

Therefore in the limit $n \to \infty$ and $\epsilon \to 0$ we have $v \geq u_\lambda^{\pi*}$. This completes the proof of part (a).

Part (b):

If $v \leq Lv$ then there exists some policy $\pi = (d, d, ..)$ such that $v \leq r_d + \lambda P_d v$. From Lemma 6.1.2 in Puterman it follows that $v \leq (I - \lambda P_d)^{-1} r_d = u_\lambda^\pi$. Therefore $v(s) \leq \max_\pi \{u_\lambda^\pi(s)\} = u_\lambda^{\pi*}$.

From parts (a) and (b) it follows that if $v = Lv$ then $v = u_\lambda^{\pi*}$.