Si assume che ogni azione sia indipendente rispetto alle azioni precedenti e siano valide le ipotesi di Markov.
Si assume che la funzione di costo $c(x, a, y)$ è strettamente positiva, cioè sempre maggiore di zero.
Si assume che $\gamma \in [0, 1]$.
Si assume che il valore di uscita sia limitato, cioè $\exists M \in \mathbb{R}^+ : \forall x \in X, \forall a \in A(x) : |c(x, a)| \leq M$.
Si assume che il valore di uscita sia limitato, cioè $\exists M \in \mathbb{R}^+ : \forall x \in X, \forall a \in A(x) : |c(x, a)| \leq M$.

Si assume che:
1. Esistano delle azioni ottime in ogni stato, per ciascun x iniziale.
2. Esistano l'operatore di programmazione dinamica T (NOTE BENE: NON SONO NECESSARIAMENTE LIMITATI).
3. Esistano una $\bar{\pi}$ ottima e un valore di partenza $\bar{x}$ tali che:
$$\sum_{t=0}^{\infty} \gamma^t \mathbb{E}_{\bar{\pi}}[c(x_t, \bar{\pi}(x_t))] < \infty$$
4. Esistano $\bar{\pi}$ parzialmente ottime, ottime, ottime $\bar{x}$ tali che:
$$\sum_{t=0}^{\infty} \gamma^t \mathbb{E}_{\bar{\pi}}[c(x_t, \bar{\pi}(x_t))] < \infty$$

Si consideri un'azione a ottima per x iniziale.
Si ha $\forall \pi \in \Pi$ e $\forall x \in X$:
$$v_\pi(x) = \sum_{t=0}^{\infty} \gamma^t \mathbb{E}_\pi[c(x_t, \pi(x_t))] \geq \sum_{t=0}^{\infty} \gamma^t \mathbb{E}_\pi[c(x_t, a)] = v_a(x)$$
$$\Rightarrow v_\pi(x) \geq v_a(x)$$
$$\Rightarrow v_\pi(x) \geq v_a(x) \geq v_{\bar{\pi}}(x)$$
$$\Rightarrow v_\pi(x) \geq v_{\bar{\pi}}(x)$$
$$\Rightarrow v_\pi(x) \geq v_{\bar{\pi}}(x)$$
$$\Rightarrow v_\pi(x) \geq v_{\bar{\pi}}(x)$$

Si ha $\forall \pi \in \Pi$ e $\forall x \in X$:
$$v_\pi(x) = \sum_{t=0}^{\infty} \gamma^t \mathbb{E}_\pi[c(x_t, \pi(x_t))] \geq \sum_{t=0}^{\infty} \gamma^t \mathbb{E}_\pi[c(x_t, a)] =