The optimal value function satisfies:

$$\forall s : V(s) = \max_a \sum_{s'} T(s,a,s') \left[R(s,a,s') + \gamma V(s') \right].$$

We can relax these non-linear equality constraints to inequality constraints:

$$\forall s : V(s) \geq \max_a \sum_{s'} T(s,a,s') \left[R(s,a,s') + \gamma V(s') \right].$$

Equivalently, ($x \geq \max_i y_i$ is equivalent to $\forall i \;\; x \geq y_i$), we have:

$$\forall s, \forall a : V(s) \geq \sum_{s'} T(s,a,s') \left[R(s,a,s') + \gamma V(s') \right]. \quad (1)$$

The relaxation still has the optimal value function as one of its solutions, but we might have introduced new solutions. So we look for an objective function that will favor the optimal value function over other solutions of (1). To this extent, let's consider the Bellman operator F, which is defined as:

$$(F(V))(s) = \max_a \sum_{s'} T(s,a,s') \left[R(s,a,s') + \gamma V(s') \right].$$

The Bellman operator F satisfies the following monotonicity property:

$$\forall s \;\; V_1(s) \geq V_2(s) \;\; \text{implies :} \;\; \forall s \;\; (F(V_1))(s) \geq (F(V_2))(s)$$

Any solution to (1) satisfies $V \geq F(V)$ (where $\geq$ is component-wise), hence also: $F(V) \geq F^2(V) = F(F(V))$, hence also: $F^2(V) \geq F^3(V)$ ... "$F^{\infty-1}$"$(V) \geq F^\infty(V) = V^*$. Stringing these together, we get for any solution V of (1) that the following holds:

$$V \geq V^*$$

Hence to find V^* as the solution to (1), it suffices to add an objective function which favors the smallest solution:

$$\min_V c^\top V \;\; \text{s.t.} \forall s, \forall a : V(s) \geq \sum_{s'} T(s,a,s') \left[R(s,a,s') + \gamma V(s') \right]. \quad (2)$$

If $c(s) > 0$ for all s, the unique solution to (2) is V^*.

Taking the Lagrange dual of (2), we obtain another interesting LP:

$$\begin{aligned} &\max_{\lambda \geq 0} \sum_{s,a,s'} T(s,a,s') \lambda(s,a) R(s,a,s') \\ &\text{s.t.} \;\; \forall s \;\; \sum_a \lambda(s,a) = \mu_0(s) + \gamma \sum_{s',a} \lambda(s',a) T(s',a,s) \end{aligned}$$

We can interpret $\lambda(s,a)$ as the discounted state-action visitation frequency for (s,a) under the initial distribution μ_0 and with discounting γ and when acting according to the policy $\pi : \Pr(a_t = a | s_t = s) = \frac{\lambda(s,a)}{\sum_b \lambda(s,b)}$. I.e., $\lambda(s,a) = \sum_{t=0}^\infty \gamma^t \Pr(s_t = s, a_t = a | \pi, s_0 \sim \mu_0)$.