$$\boxed{Q_\pi^{(t-1)}(s,\pi(s)) = \sum_{s' \in \text{ States}} T(s,\pi(s),s') \Big[\text{Reward}(s,\pi(s),s') + \gamma V_\pi^{(t-1)}(s')\Big]}$$

Remarque : en notant S le nombre d'états, A le nombre d'actions par états, S' le nombre de successeurs et T le nombre d'itérations, la complexité en temps est alors de $\mathcal{O}(T_{PE}SS')$.

❒ **Q-value optimale** – La Q-value optimale $Q_{\text{opt}}(s,a)$ d'un état s avec l'action a est définie comme étant la Q-value maximale atteinte avec n'importe quelle politique. Elle est calculée avec la formule :

$$\boxed{Q_{\text{opt}}(s,a) = \sum_{s' \in \text{ States}} T(s,a,s') \Big[\text{Reward}(s,a,s') + \gamma V_{\text{opt}}(s')\Big]}$$

❒ **Valeur optimale** – La valeur optimale $V_{\text{opt}}(s)$ d'un état s est définie comme étant la valeur maximum atteinte par n'importe quelle politique. Elle est calculée avec la formule :

$$\boxed{V_{\text{opt}}(s) = \max_{a \in \text{ Actions}(s)} Q_{\text{opt}}(s,a)}$$

❒ **Politique optimale** – La politique optimale π_{opt} est définie comme étant la politique liée aux valeurs optimales. Elle est définie par :

$$\forall s, \quad \boxed{\pi_{\text{opt}}(s) = \underset{a \in \text{ Actions}(s)}{\text{argmax}} Q_{\text{opt}}(s,a)}$$

❒ **Itération sur la valeur** – L'algorithme d'itération sur la valeur (en anglais *value iteration*) vise à trouver la valeur optimale V_{opt} ainsi que la politique optimale π_{opt} en deux temps :

— Initialisation : pour tout état s, on a

$$\boxed{V_{\text{opt}}^{(0)}(s) \longleftarrow 0}$$

— Itération : pour t allant de 1 à T_{VI}, on a

$$\forall s, \quad \boxed{V_{\text{opt}}^{(t)}(s) \longleftarrow \max_{a \in \text{ Actions}(s)} Q_{\text{opt}}^{(t-1)}(s,a)}$$

avec

$$\boxed{Q_{\text{opt}}^{(t-1)}(s,a) = \sum_{s' \in \text{ States}} T(s,a,s') \Big[\text{Reward}(s,a,s') + \gamma V_{\text{opt}}^{(t-1)}(s')\Big]}$$

Remarque : si $\gamma < 1$ ou si le graphe associé au processus de décision markovien est acyclique, alors l'algorithme d'itération sur la valeur est garanti de converger vers la bonne solution.