Algorithm 1 Value Iteration

Initialization:

Require: $V(s) \in R$ and $\pi(s) \in A(s)$ for all $s \in S$

loop: $\Delta \leftarrow 0$

  loop: for each $s \in S$

    $v \leftarrow V(s)$

    $V(s) \leftarrow max_a \sum_{s',r} p(s', r|s, a)[r + \gamma V(s')]$

    $\Delta \leftarrow max(\Delta, |v - V(s)|)$

  end loop

  untill $\Delta < \theta$ (with θ as the convergence criteria)

end loop

return $\pi \approx \pi_*$, such that $\pi(s) = argmax_a \sum_{s',r} p(s', r|s, a)[r + \gamma V(s')]$