Definition of P:
$P(s,a,s') = \Pr(S_{t+1} = s', | S_t = s, A_t = a)$

Definition of $d_0(s)$:
$d_0(s) = \Pr(S_0 = s)$

Definition of π:
$\pi(s,a) = \Pr(A_t = a | S_t = s)$

Definition of R:
$R(s,a,s') = \mathbf{E}[R_t | S_t = s, A_t = a, S_{t+1} = s']$

Definition of v^π:
$v^\pi(s) = \mathbf{E}[\sum_{k=0}^{\infty} \gamma^k R_{t+k} | S_t = s, \pi]$

Definition of J:
$J(\pi) = E[\sum_{t=0}^{\infty} \gamma^t R_t | \pi]$

Definition of q^π:
$q^\pi(s,a) = \mathbf{E}[\sum_{k=0}^{\infty} \gamma^k R_{t+k} | S_t = s, A_t = a, \pi]$

Bellman Operator T:

$$Tv(s) = \max_a \sum_{s'} P(s,a,s')(R(s,a,s') + \gamma v(s'))$$
$$Tq(s,a) = \sum_{s'} P(s,a,s') \left(R(s,a,s') + \gamma \max_{a'} q(s',a') \right)$$

Definition of G:
$\sum_{t=0}^{\infty} \gamma^t R_t$

Definition of G_t:
$\sum_{k=0}^{\infty} \gamma^k R_{t+k}$

Equation Relating J and G:
$J(\pi) = \mathbf{E}[G | \pi]$

The Bellman Equation for v and q:

$$
\begin{aligned}
v^\pi(s) &= \sum_a \pi(s,a) \sum_{s'} P(s,a,s')(R(s,a,s') + \gamma v^\pi(s')) \\
&= \sum_a \pi(s,a) \sum_{s'} P(s,a,s')(R(s,a,s') + \gamma \sum_{a'} \pi(s',a') q^\pi(s',a')) \\
q^\pi(s,a) &= \sum_{s'} P(s,a,s')(R(s,a,s') + \gamma v^\pi(s')) \\
&= \sum_{s'} P(s,a,s')(R(s,a,s') + \gamma \sum_{a'} \pi(s',a') q^\pi(s',a'))
\end{aligned}
$$

The Bellman Optimality Equations for v and q:

$$
\begin{aligned}
v^\star(s) = v^{\pi^\star}(s) &= \max_a \sum_{s'} P(s,a,s')(R(s,a,s') + \gamma v^\star(s')) \\
q^\star(s,a) = q^{\pi^\star}(s,a) &= \sum_{s'} P(s,a,s') \left(R(s,a,s') + \gamma \max_{a'} q^*(s',a') \right)
\end{aligned}
$$

TD Update:

$$
\begin{aligned}
\delta &= r + \gamma v(s') - v(s) \\
&or \\
\delta_t &= R_t + \gamma v(S_{t+1}) - v(S_t) \\
\text{tabular:} \\
v(s) &= v(s) + \alpha\delta \\
&or \\
v(S_t) &= v(S_t) + \alpha\delta_t \\
linear: w &= w + \alpha\delta_t \phi(S_t) \\
general: w &= w + \alpha\delta_t \frac{\partial v_w(S_t)}{\partial w}
\end{aligned}
$$

Sarsa and Q-Learning Update:

$$
\begin{aligned}
&Sarsa: \\
&\delta_t = R_t + \gamma q(S_{t+1}, A_{t+1}) - q(S_t, A_t) \\
&Q-Learning: \\
&\delta_t = R_t + \gamma \max_{a'} q(S_{t+1}, a') - q(S_t, A_t) \\
&Both: \\
&q(S_t, A_t) = q(S_t, A_t) + \alpha\delta_t
\end{aligned}
$$

TD(λ):

$$\begin{aligned}
\delta_t = R_t + v(S_{t+1}) - v(S_t)\\
\forall s \in S : e(s) = \gamma\lambda e(s)\\
e(S_t) = e(S_t) + 1\\
\forall s \in S : v(s) = v(s) + \alpha\delta_t e(s)\\
Sarsa(\lambda) :\\
\delta_t = R_t + \alpha q(S_{t+1}, A_{t+1}) - q(S_{t+1}, A_{t+1})\\
\forall s \in S, a \in A : e(s, a) = \gamma\lambda e(s, a)\\
e(S_t, A_t) = e(S_t, A_t) + 1\\
\forall s \in S, a \in A : q(s, a) = q(s, a) + \alpha\delta_t e(s, a)\\
Q(\lambda) :\\
\delta_t = R_t + \alpha \max_{a'} q(S_{t+1}, a') - q(S_{t+1}, A_{t+1})\\
\forall s \in S, a \in A : e(s, a) = \gamma\lambda e(s, a)\\
e(S_t, A_t) = e(S_t, A_t) + 1\\
\forall s \in S, a \in A : q(s, a) = q(s, a) + \alpha\delta_t e(s, a)
\end{aligned}$$

Policy Gradient Formulae (both formulae and the supporting term):
$\partial J/\partial\theta = \sum_s d^\pi(s) \sum_a q^\pi(s, a) \frac{\partial\pi(s,a,\theta)}{\partial\theta}$
$\partial J/\partial\theta = \sum_s d^\pi(s) \sum_a \pi(s, a, \theta) * q^\pi(s, a) \frac{\partial \ln \pi(s,a,\theta)}{\partial\theta}$
$d^\pi(s) = \sum_{t=0}^{\infty} \gamma^t * \Pr(S_t = s|\pi)$