Dynamic Programming **Dynamics:**
- $x_{k+1} = f_k(x_k, u_k, w_k),\ k = 0, 1, ..., N-1$

where $x_k \in S_k, u_k \in U_k(x_k)$, and $w_k \backsim p_{w_k|x_k,u_k}$ with $p_{w_k|x_k,u_k,*} = p_{w_k|x_k,u_k}, \forall * \in \{x_l, u_l, w_l | l < k\}$
- admissible policy:$\pi = (\mu_0(.), \mu_1(.), ..., \mu_{N-1}(.))$

$u_k = \mu_k(x_k), u_k \in U_k(x_k), k = 0, 1, ..., N-1$

Expected Cost:
Given $x \in S_0$, the expected closed loop cost of starting at $x_0 = x$ associated with policy π is: $J_\pi(x) = \underset{(X_1, W_0|x_0=x)}{E}[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k)]$, where $X_1 = (x_1, ..., x_N), W_0 = (w_0, ..., w_{N-1})$

Objective:
Construct an optimal policy π^* s.t. $\forall x \in S_0$:

$$\pi^* = \underset{\pi \in \Pi}{argmin} J_\pi(x)$$

*Open loop control can never give better performance than closed loop control (u_k depends on x_k) since open loop control is a special case of closed loop control. In the absence of disturbances w_k, the two give theoretically the same performance.
Consider a system with N_x distinct states and N_u distinct control inputs: There are a total of N_u^N different open loop strategies. There are a total of $N_u(N_u^{N_x})^{N-1}$ different closed loop strategies.

Transition Probability:
$P_{ij}(u, k) = P(x_{k+1} = j | x_k = i, u_k = u) = p_{x_{k+1}|x_k,u_k}(j|i, u) = p_{w_k|x_k,u_k}(j|i, u) = \sum_{\bar{w}_k | f_k(i,u,\bar{w}_k)=j} p_{w_k|x_k,u_k}(\bar{w}_k|i, u)$

Principle of Optimality: Let π^* be an optimal policy. Consider the subproblem whereby we are at $x \in S_i$ at time i and we want to minimize: $\underset{X_{i+1}, W_i | x_i = x}{E}[g_N(x_N) + \sum_{k=i}^{N-1} g_k(x_k, \mu_k(x_k), w_k)]$
where $X_{i+1} = (x_{i+1}, ..., x_N) and W_i = (w_i, ..., w_{N-1})$.
Then the truncated policy $\pi^* = (\mu_i^*(.), ..., \mu_{N-1}^*(.))$ is optimal for this problem

DPA:
Initialization: $J_N(x) = g_N(x), \forall x \in S_N$
Recursion:
$J_k(x) = \underset{u \in U_k(x)}{min} \underset{(w_k|x_k=x, u_k=u)}{E}[g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k))], \forall x \in S_k, k = N-1, ..., 0$
*We calculate cost-to-go J_k with expected value, where we don't consider variance $Var(x) = E(x^2) - E(x)^2$
*Computation: $N_u N_x (N-1) + N_u$ operations
Time Lags: Assume the dynamics becomes:

$$x_{k+1} = f_k(x_k, x_{k-1}, u_k, u_{k-1}, w_k)$$

Let $y_k = x_{k-1}, s_k = u_{k-1}, \tilde{x}_k = (x_k, y_k, s_k)$
$\tilde{x}_{k+1} = \begin{bmatrix} x_{k+1} \\ y_{k+1} \\ s_{k+1} \end{bmatrix} = \begin{bmatrix} f_k(x_k, y_k, u_k, s_k, w_k) \\ x_k \\ u_k \end{bmatrix} = \tilde{f}_k(\tilde{x}_k, u_k, w_k)$

Correlated Disturbances
If $w_k = C_k y_{k+1}, y_{k+1} = A_k y_k + \xi_k$, where $\xi_k, k = 0, ..., N-1$ are independent random variables.
- Let the augmented state vector $\tilde{x}_k = (x_k, y_k)$. Note that now y_k must be observed at time k, which can be done using a state estimator.
- $\tilde{x}_{k+1} = \begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} f_k(x_k, u_k, w_k = C_k(A_k y_k + \xi_k)) \\ A_k y_k + \xi_k \end{bmatrix} = \tilde{f}_k(\tilde{x}_k, u_k, \xi_k)$

Forecast
At the beginning of each period k, we receive a prediction y_k (forecast), we know a collection of distributions $\{p_{w_k|y_k}(.|.), ...\}$ and priori The forecast itself has a given a-priori probability distribution $p(\xi_k)$ with $y_{k+1} = \xi_k$. y_{k+1}: this event happens on day k+1, ξ_k: the forecast about y_{k+1} on day k
- New state vector: $\tilde{x}_k = (x_k, y_k)$, new disturbance: $\tilde{w}_k = (w_k, \xi_k)$

$\tilde{x}_{k+1} = \begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} f_k(x_k, u_k, w_k) \\ \xi_k \end{bmatrix} = \tilde{f}_k(\tilde{x}_k, u_k, \xi_k)$ •
The dynamics becomes: $J_k(\tilde{x}) = \underset{u \in U_k(x)}{min} \underset{(w_k|y_k=y)}{E}[g_k(x, u, w_k) + \underset{\xi_k}{E}[J_{k+1}(f_k(x, u, w_k), \xi_k)]]$
$= \underset{u \in U_k(x)}{min} \underset{(w_k|y_k=y)}{E}[g_k(x, u, w_k) + \sum_{i=1}^m p_{\xi_k}(i) J_{k+1}(f_k(x, u, w_k), i)]$
$\forall x \in S_k, y \in \{1, ..., m\}, k = N-1, ..., 0$

Infinite Horizon Problem: as N goes infinity, let $\mathbf{V_1}(.) = \mathbf{J_{N-1}}(.)$, V converges, so we have $J(x) = \underset{u \in U_k(x)}{min} \underset{(w|x=x, u=u)}{E}[g(x, u, w) + J(f(x, u, w))], \forall x \in S$, i.e. **Bellman Equation** $->$ optimal policy is time invariant.

Stochastic Shortest Path Problem
- **Dynamics:**

$x_{k+1} = w_k, P(w_k = j | x_k = i, u_k = u) = P_{ij}(u)$(time-invariant transition probability), $\forall x_k \in S, u \in U(i)$
U, S are finite,
- **Cost:**

$J_\pi(i) = \underset{(X_1, W_0|x_0=i)}{E}[\sum_{i=0}^{N-1} g(x_k, \mu_k(x_k), w_k)]$

Assumption 4.1 Cost-free termination state:
State 0 is denoted as the termination state with $S = 0, 1, ..., n$, where
$P_{00}(u) = 1, g(0, u, 0) = 0, \forall u \in U(0)$
A stationary policy μ is said to be **proper** if, when using this policy, there exists an integer m such that: $P(x_m = 0 | x_0 = i) > 0$

Assumption 4.2 proper policy: There exists at least one proper policy $\mu \in \Pi$. Furthermore, for every improper policy $\mu^{'}$, the corresponding cost function $J_{\mu^{'}}(i)$ is infinity for at least one state $i \in S$.
*This assumption is required in order to guarantee that a unique solution to the BE exists for the SSP problem, which will then be the optimal cost.
*It ensures that a policy exists for which the probability of reaching the termination state goes to one as the time horizon N goes to infinity. It also ensures that the policies for which this does not occur incur infinite cost, which ensures that there are no non-positive cycles.

Theorem for SSP:*under assumption 4.1 and 4.2,
1.Given any initial conditions $V_0(1), ..., V_0(n)$, the sequence $V_l(i)$:
$V_{l+1}(i) = \underset{u \in U(i)}{min}(q(i, u) + \sum_{j=1}^n P_{ij}(u) V_l(j)), \forall i \in S^+ (1)$
where $S^+ = S \backslash 0$ and $q(i, u) = \underset{(w|x=i, u=u)}{E}[g(x, u, w)]$
converges to the optimal cost $J^*(i)$ for all $i \in S^+$
2.The optimal cost satisfy the BE:
$J^*(i) = \underset{u \in U(i)}{min}(q(i, u) + \sum_{j=1}^n P_{ij}(u) J^*(j)), \forall i \in S^+$
3. The solution to the BE is unique
4. The minimizing u for each $i \in S^+$ of the BE gives an optimal policy, which is proper.

Value Iteration:(1) above, until a threshold for $\|V_{l+1}(i) - V_l(i)\|$ is reached

Policy Iteration: • **initialization:** Initialize with a proper policy $\mu^0 \in \Pi$
- **Policy evaluation:** Given a policy μ_h, solve for the corresponding cost J_{μ^h} by solving the linear system

$J_{\mu^h}(i) = q(i, \mu^h(i)) + \sum_{j=1}^n P_{ij}(\mu^h(i)) J_{\mu^h}(j), \forall i \in S^+$
- **Policy Improvement:** Obtain a new stationary policy μ^{h+1}:

$\mu^{h+1}(i) = \underset{u \in U(i)}{argmin}(q(i, u) + \sum_{j=1}^n P_{ij}(u) J_{\mu^h}(j)), \forall i \in S^+$, iterate until $J_{\mu^{h+1}}(i) = J_{\mu^h}(i)$ for all $i \in S^+$