Unit Commitment

1. **States**(S_t)
   (a) Forecast net load (d^f_{t+1}) [Belief state]
   (b) Sequence of start-up status $[v_{it}, v^{old}_{it}]$ of ith generator with a length of UT_i [Physical state]
   (c) Sequence of shut-down status $[w_{it}, w^{old}_{it}]$ of ith generator with a length of DT_i [Physical state]
   (d) Power output p_{it} at ith generator
   (e) Current time step t [Physical State]
2. **Actions** (A_t)
   (a) On/off status u_{it+1} at ith generator
   (b) Power output p_{it+1} at ith generator
3. **Transition function**
   (a) Forecast net load
   $$d^f_{t+2} = \text{forecast}(d^f_{t+1})$$
   (b) On/off status
   $$u_{it+1} = \text{MustOn\&Off}(u_{it+1})$$
   (c) Start-up status
   $$v_{it+1} = \max(0, u_{it+1} - u_{it})$$
   (d) Shut-down status
   $$w_{it+1} = -\min(0, u_{it+1} - u_{it})$$
   (e) Sequence of Start-up status
   $$\mathbf{v}_{it+1} = \left[v_{it+1}, v^{old}_{it}[1:]\right]$$
   (f) Sequence of Shut-down status
   $$\mathbf{w}_{it+1} = \left[w_{it+1}, w^{old}_{it}[1:]\right]$$

(g) Power output

$$p_{it+1} = u_{it+1}\text{clip}(p_{it+1})$$

4. **Action Correction, Violation Cost and Reward**

(a) Minimum Up-time & Down-time Constraints are enforced by MustOn&Off($\cdot$) which uses the inequality $\sum_k^{UT} v(k) \leq u(t+1)$ and $\sum_k^{DT} w(k) \leq 1 - u(t+1)$ to find out the generators that must be kept on & off. The violation results in a cost, but does not directly lead to "truncated" directly.

(b) Generation Limit Constraints are enforeced by a clip function and the corrected on/off status.

(c) Ramping Constraints are evaluated through $p(t{+}1){-}p(t) \leq RUu(t){+}SUv(t+1)$ and $p(t) - p(t+1) \leq RDu(t+1) + SDw(t+1)$. The violation results in a cost and "truncated" without good correction method.

(d) Flow Balance Constraints are relaxed through slacks that represent overflow and underflow. The energy waste or unfulfillment is then explicitly computed using a load shedding cost in the reward.

(e) Reserve Requirement Constraints are relaxed through slacks that represent the unfulfillment $r_i + s \geq R_i$. The intermediate variable r_i represents the reserve at ith generator. To compute r_i, first evaluate $\min(u(t{+}1)P_{max}{-}p(t{+}1), RUu(t){+}SUv(t{+}1){+}p(t){-}p(t{+}1)))$, then evaluate $\max(\cdot, 0)$. This finds the best reserve that can be achieved without violating both power limit and ramping-up constraints. The unfulfillment is then explicitly computed using a reserve penalty cost in the reward.