Algorithm 1 PPO

for $iteration = 1, 2, \ldots$ **do**
  for $actor = 1, 2, \ldots, N$ **do**
    Run policy $\pi_{\theta_{old}}$ in environment for T time steps
    Compute advantage estimates $\hat{A}_1, \ldots, \hat{A}_T$
  end for
  Optimize surrogate L wrt. θ, with K epochs and minibatch size $M \leq NT$
  $\theta_{old} \leftarrow \theta$
end for

Algorithm 2 Game Theory Controller

for Every time step **do**
    Calculate target seeking command $\mathbf{x}_{tsCmd}$ (Eq.: 3.12)
    for All map measurements from $\mathbf{x}_{Map}$ **do**
        Denormalize measurement (Eq.: 3.14)
        Add margin of safety (Eq.: 3.15)
        Calculate altitude difference $\Delta h_{ObsSafe_j}$ to aircraft (Eq.: 3.16)
        if $\Delta h_{ObsSafe_j} > 0$ **then**
            Add measurement to set of critical measurements $\mathcal{M}_{crit}$ (Eq.: 3.17)
        end if
    end for
    for All measurements in $\mathcal{M}_{crit}$ **do**
        Calculate local obstacle avoidance vector (Eq.: 3.20)
    end for
    Sum over all local avoidance vectors (Eq.: 3.22)
    Transform to global coordinate frame to receive $\mathbf{x}_{oaCmd}$ (Eq.: 3.23)
    Calculate obstacle avoidance weight w_{oa} based on critical zone weight (Eq.: 3.24)
    Calculate target seeking weight w_{ts} as $1 - w_{oa}$ (Eq.: 3.13)
    Calculate command vector $\mathbf{x}_{HSaCmd} = w_{oa}\mathbf{x}_{oaCmd} + w_{ts}\mathbf{x}_{tsCmd}$ (Eq.: 3.11)
end for