COS 511: Theoretical Machine Learning

Hint on HW#8, problem 2 Spring, 2019

Find an appropriate auxiliary function for this algorithm, and then apply the proof technique given in class. Keep in mind that the properties of an auxiliary function $A(\mathbf{p})$ must hold for *any* distribution $\mathbf{p}$, not just the distributions computed by a particular algorithm.

Also, at some point in your proof, you might find it helpful to argue that, for each $j \in \{1, \ldots, n\}$, there exists a continuous, fixed function ϕ_j mapping distributions to distributions such that, on each round t, if $j = (t \bmod n)$, then $\mathbf{p}_{t+1} = \phi_j(\mathbf{p}_t)$. (Here, "fixed" means that, aside from taking $\mathbf{p}_t$ as an argument, the function ϕ_j itself does not depend on any of the quantities that vary as our iterative algorithm is run, such as $\boldsymbol{\lambda}_t$; however, it still might depend on quantities that do not vary but are fixed at the start of the algorithm, such as the x_i's and f_j's.)

Finally, you can take as a given fact that if a sequence $\mathbf{z}_1, \mathbf{z}_2, \ldots$ can be partitioned into finitely many subsequences (so that each $\mathbf{z}_t$ appears in one subsequence), and if each of these converges to the same point $\mathbf{z}^*$, then the entire sequence of $\mathbf{z}_t$'s also converges to $\mathbf{z}^*$. (This can be proved directly using the definition of convergence.)