6.7950 Fall 2022: - Recitation 5 Handout

The end result of today's recitation is to prove the convergence of a simplified version of stochastic value iteration. It's a broad result that can be applied to other optimization algorithms, like stochastic gradient descent. A very similar result is going to be seen in future lectures, so the intent of today's recitation is to provide an initial expsoure to it and facilitate the lectures. For this reason, this handout is provided with solutions before the actual recitation.

1 Supermartingales

Definitions

- Consider a stochastic process $\{Z_t,\ t \geq 1\}$ with $\mathbb{E}[Z_t]$ bounded. Then it will be

  Supermartingale if $\mathbb{E}[Z_{t+1}|Z_t, \dots, Z_1] \leq Z_t$
  Martingale if $\mathbb{E}[Z_{t+1}|Z_t, \dots, Z_1] = Z_t$
  Submartingale if $\mathbb{E}[Z_{t+1}|Z_t, \dots, Z_1] \geq Z_t$

- The history of the process up until point t appears very often, so we will call it the filtration F_t of the process up until time t.

  $$F_t = [Z_t, Z_{t-1}, \dots, Z_2, Z_1]$$

  The notion of a filtration is more rigorously defined in the literature, but for our purposes it's not necessary to be very precise. You can think of each F_t contains all known information at time t and no information is lost from t to $t+1$, only gained.

Consider the following questions to help you build intuition about these definitions.

1. Are martingale processes also Markovian?
2. Given random i.i.d. variables $X_i \geq 0$ with $\mathbb{E}[X_i] = 1$, prove that $Z_t = \prod_1^t X_i$ is a Martingale.

Solution:

1. No, the Martingale property is a property of a conditional expectation, while the Markov property is associated with the whole conditional expectation. As a counter example, consider the process

   $$Z_{t+1} = Z_t + Z_0 w_t, \text{ with i.i.d. } w_t \sim \mathcal{N}(0,1) \text{ and } Z_0 = w_0$$

We can verify that $\mathbb{E}[Z_{t+1}|F_t] = Z_t$, so it's Martingale, but it's not Markovian as $p(Z_{t+1}|F_t) = \mathcal{N}(Z_t, Z_0^2)$ but $p(Z_{t+1}|Z_t)$ will not be a gaussian distribution because it involves the product of two independent gaussians plus a Z_t term.

2. We can verify the Martingale property using basic properties of expectations

$$\mathbb{E}\left[Z_{t+1}|F_t\right] = \mathbb{E}\left[\prod_1^{t+1} X_i|F_t\right] = \mathbb{E}\left[X_{t+1}\prod_1^{t} X_i|F_t\right] = \mathbb{E}\left[X_{t+1}|F_t\right]\prod_1^{t} X_i = 1.\prod_1^{t} X_i = Z_t$$

Supermartingale convergence theorem

The supermartingale convergence theorem can be seen as a more generalized version of proving the convergence of a bounded monotonically increasing sequence. Let's first present the theorem and then explore some examples to better understand it.

Consider the scalar random variables X_t, Y_t, Z_t for $k = 0, 1, \ldots$ and the filtration F_t involving the history of all these variables up to time t, that is

$$F_t = (X_t, Y_t, Z_t, \ldots, X_2, Y_2, Z_2, X_1, Y_1, Z_1)$$

If the following conditions hold for all t.

- X_t, Y_t, Z_t are all nonnegative
- $\sum Z_t < \infty$
- $\mathbb{E}[Y_{t+1}|F_t] \leq Y_t - X_t + Z_t$

then the theorem states that we have with probability 1 that

- $\lim_{t\to\infty} Y_t$ exists and is finite
- $\sum X_t < \infty$

We can notice that the third condition looks "almost supermartingale", but Z_t and X_t provide some limited slack to much this tendency can be violated. Let's look at some special cases to further improve our intuition.

1. When $Z_t = 0$ and the other conditions of the theorem holds, show that $\mathbb{E}[Y_{t+1}] \leq \mathbb{E}[Y_t]$.
2. When $Z_t = 0$ and the other conditions of the theorem holds, show that having $X_t = 1$ would generate a contradiction without using the result of the theorem.

Solution:

1. Applying $Z_t = 0$ in the condition of the theorem we get that

$$\mathbb{E}[Y_{t+1}|F_t] \leq Y_t - X_t \leq Y_t$$

By applying the expectation operator to both sides we can use the law of total expectation (recall that is can be stated as $\mathbb{E}[\mathbb{E}[a|b]] = \mathbb{E}[a]$) to get

$$\mathbb{E}[\mathbb{E}[Y_{t+1}|F_t]] = \mathbb{E}[Y_{t+1}] \leq E[Y_t]$$

2. By applying the expectation operator on both sides of the third condition, we get that $\mathbb{E}[Y_{t+1}] \leq \mathbb{E}[Y_t] - 1$, so the mean of $\mathbb{E}[Y_t]$ decreases at least linearly, so eventually it will become negative. However, that contradicts the first condition as Y_t are all nonnegative and therefore $\mathbb{E}[Y_t] \geq 0$.

2 Convergence of (simplified) value iteration

We can use the previous convergence theorem to prove the convergence of a simplified stochastic version of value iteration, but we are going to present in a more generic way so it's applicable to other algorithms too. Specifically, consider the following update rule

$$x_{t+1} = x_t + \eta_t g(x_t, w_t)$$

We can think of x as the variable we want to update (say, the value function) using the deterministic learning rate η_t. Furthermore, $g(x_t, w_t)$ is an update function that is corrupted by some independent random disturbance w_t. This update function could be, for example, the temporal difference corrupted by some noise $g(x_t, w_t) = \delta(x_t) + w_t$ or it could be a noisy gradient descent direction with respect to some loss f as in $g(x_t, w_t) = -\nabla f_{x_t} + w_t$. Let this optimal be realized at x^*.

1. Assume there is positive scalar c and nonnegative scalars K_1, K_2 for which the following assumptions hold:

   Pseudo-gradient property $(x^* - x)^\top \mathbb{E}_w[g(x, w)] \geq c \left\| x^* - x \right\|_2^2$

   Intuitively, we can think of this property as describing the notion that, on average, the update direction g should have a component pointing in the direction that takes x to the optimal x^*. If we had for example $g(x, w) = x^* - x$, then the property would hold with $c = 1$ and the first update would be lead to the optimal value, while the opposite direction $g(x, w) = -(x^* - x)$ would not satisfy this property and would not lead to the desired x. It should also be intuitive that $c = 0$ would also be undesirable

   Bounded variance $\mathbb{E}_w[\left\| g(x, w) \right\|_2^2] \leq K_1 + K_2 \left\| x^* - x \right\|_2^2$

   As there's noise in the update direction, we assume that its variance is bounded

   Robbins-Monro step size $\sum \eta_t = \infty, \; \sum \eta_t^2 < \infty$

   A standard condition already seen in lecture, but that will make more mathematical sense once the solution is derived.

Notice that the expectation E_w in the previous assumptions is only with respect to w and not x. Then, prove that the presented iteration scheme converges to the optimal (that is, $x \to x^*$) almost surely, using the supermartingale convergence theorem.

Solution idea: Let the L_2 distance of x to the optimal at time t be $\Delta_t = \|x^* - x\|_2^2$. We want to show that it converges to 0. So we first want to show that $E[\Delta_{t+1}|F_t] = Y_t - X_t + Z_t$ for some X_t, Y_t, Z_t satisfying the conditions of the theorem. The consequences of the theorem can then be manipulated to obtain the desired result (the theorem result would only state that Δ_t converges, but not necessarily to 0).

Solution: We start by expanding the expression for $\Delta_{t+1} = \|x^* - x\|_2^2$

$$\Delta_{t+1} = \|x^* - \eta_t g(x_t, w_t) - x_t\|_2^2 = \Delta_t - 2\eta_t(x^* - x_t)^\top g(x_t, w_t) + \eta_t^2 \|g(x_t, w_t)\|_2^2$$

We can apply a conditional expectation operator to the previous equation followed by a straightforward application of the provided assumptions in order to obtain the desired condition for the supermartingale theorem.

$$\begin{aligned}\mathbb{E}[\Delta_{t+1}|F_t] &= \Delta_t - 2\eta_t\mathbb{E}[(x^* - x_t)^\top g(x_t, w_t)|F_t] + \eta_t^2\mathbb{E}[||g(x_t, w_t)||_2^2|F_t] \\ &\leq \Delta_t - 2\eta_t c\Delta_t + \eta_t^2(K_1 + K_2\Delta_t) \\ &= \underbrace{\Delta_t}_{=Y_t} - \underbrace{\Delta_t(2\eta_t c - \eta_t^2 K_2)}_{=X_t} + \underbrace{\eta_t^2 K_1}_{=Z_t}.\end{aligned}$$

Clearly, the candidates for Y_t and Z_t are not negative, but X_t could be negative for some t. As $\eta \to 0$ (which is true from $\sum \eta_t^2 < \infty$), then at some time T we are going to have $2\eta_t c \geq \eta_t^2 K_2, \forall t \geq T$ because a quadratic term shrinks faster than a linear one. So for $t \geq T$, all the candidates for X_t, Y_t and Z_t are nonnegative, which is one of the required assumptions for the theorem. The condition that $\sum Z_t < \infty$ also follows from Robbins-Monro step size because $\sum Z_t = K_1 \sum \eta_t^2 < \infty$.

Thus, we have satisfied the requirements of the supermartingale convergence theorem and can conclude that Δ_t converges a.s. to *some* value and that $\sum X_t < \infty$. In other words we have

$$\sum X_t = \sum \Delta_t 2\eta_t c - \Delta_t \eta_t^2 K_2 < \infty$$

and since Δ_t converges and $\sum \eta_t^2 < \infty$, then $\sum \Delta_t \eta_t^2 K_2 < \infty$. This wouldn't be possible if $\sum \Delta_t \eta_t c$ was unbounded, so we conclude that

$$\sum \Delta_t \eta_t c < \infty$$

However, the other condition of the Robbins-Monro step size requires that $\sum \eta_t = \infty$, so it must be that $\Delta_t \to 0$ almost surely. Finally, this means that $x \to x^*$ with probability one.