Stochastic Approximation (SA) and Monte Carlo Estimation (MCE)

methods in Simulation

- Consider the computation of the fixed point of a mapping $F : R^n \to R^n$ that is a contraction w.r.t. some norm and involves an expectation:

$$x = F(x)$$

where $F(x) = \underset{w}{E}\{f(x, w)\}$.

- For simplicity assume the random variable w takes values in a finite set W with distribution $\xi = \{\xi(w) \mid w \in W\}$.

- Now consider the fixed point iteration to solve $x = F(x)$

$$x_{k+1} = F(x_k) = E\{f(x_k, w)\}, \quad k = 1, 2, \cdots \quad \text{— (19)}$$

- Generate a sequence $\{w_1, w_2, \ldots,\}$ of samples such that the emperical frequency of each $w \in W$ is equal to its probability, i.e.,

$$\xi(w) = \lim_{k \to \infty} \frac{1}{k} \sum_{t=1}^{k} \delta(w_t = w), \quad \forall w \in W.$$

Monte Carlo Estimation (MCE) method

Consider simulation-based approximation of the iteration (19)

$$x_{k+1} = \frac{1}{k} \sum_{t=1}^{k} f(x_k, w_t), \quad k = 1, 2, \cdots. \qquad \text{—— (20)}$$

The MCE method has a major flaw: it requires, for each k, the computation of $f(x_k, w_t)$ for all sample values w_t, $t = 1, \cdots, k$.

Stochastic Approximation (SA) method

Instead, consider a similar iteration

$$x_{k+1} = \frac{1}{k}\sum_{t=1}^{k} f(x_t, w_t), \quad k=1,2,\cdots\cdots \qquad (21)$$

which requires much less computation because it needs only one value of f per sample w_t. It can also be written as

$$x_{k+1} = (1-\gamma_k)x_k + \gamma_k f(x_k, w_k), \; k=1,2,\cdots \qquad (22)$$

with the stepsize $\gamma_k = 1/k$.

why?

$k=1$: $x_2 = f(x_1, w_1)$

$k=2$: $x_3 = \dfrac{f(x_1,w_1)+f(x_2,w_2)}{2} = \dfrac{1}{2}x_2 + \dfrac{1}{2}f(x_2,w_2)$

$k=3$: $x_4 = \dfrac{f(x_1,w_1)+f(x_2,w_2)+f(x_3,w_3)}{3} = \dfrac{f(x_1,w_1)+f(x_2,w_2)}{3} + \dfrac{1}{3}f(x_3,w_3)$

$= \dfrac{2}{3}x_3 + \dfrac{1}{3}f(x_3,w_3)$

$\vdots$

- Under mild conditions, one can show that the limit $\bar{x} = \lim_{k\to\infty} x_k$ satisfies $\bar{x} = F(\bar{x}) = \underset{w}{E}\{f(\bar{x}, w)\}$.
- More general stepsize rules for this convergence are

$$0 < \gamma_k \le 1, \; \forall k \;\;\&\;\; \lim_{k\to\infty}\gamma_k = 0 \;\;\&\;\; \sum_{k=1}^{\infty}\gamma_k = \infty \qquad (23)$$

e.g.) $\gamma_k = \dfrac{c_1}{k+c_2}$: where positive constants c_1 & c_2 are chosen to satisfy (23).

Projected (Bellman) Equation Methods for Finite-State MDPs

- Assume linear approximation architecture for $J_\mu(i)$

$$\tilde{J}_\mu(i;r) = \phi(i)'r$$

i.e., we want to approximate $J_\mu(i)$ within $S = \{\Phi r \mid r \in R^s\}$, the subspace spanned by s basis functions, the columns of Φ.

Assumption 1.

> The Markov chain corresponding to μ has unique steady-state probabilities $\xi_1, \cdots, \xi_n$, which are positive, i.e., under μ, we have for all $i = 1, \cdots, n$,
>
> $$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} P(i_k = j \mid i_0 = i) = \xi_j > 0, \quad j = 1, \cdots, n.$$

- Note that Assumption 1 is equivalent to assuming that the Markov chain is irreducible, i.e., has a single recurrent class and no transient states.

Assumption 2.

> The matrix Φ has rank s.

- Note that Assumption 2 is equivalent to the basis functions (the columns of Φ) being linearly independent, i.e., every vector in S is represented in the form of Φr with a unique vector r.

The Projected Bellman Equation

$$\Phi r = \Pi T_\mu(\Phi r) \longrightarrow \tilde{J}_\mu = \Pi T_\mu \tilde{J}_\mu$$

where $T_\mu(\Phi r)(i) = \sum_{j=1}^{n} P_{ij}(\mu(i))\big(g(i,\mu(i),j) + \alpha(\Phi r)(j)\big)$, $i = 1, \cdots, n$

and

$$r^* = \arg\min_{r \in R^s} \| T_\mu(\Phi r) - \Phi r \|_2^{\nu\,2} \longrightarrow \| T_\mu \tilde{J}_\mu - \tilde{J}_\mu \|_2^{\nu\,2}$$

where $\|\cdot\|_2^{\nu}$ is a weighted Euclidean (L2) norm on $\mathbb{R}^n$ of the form

$$\|\Phi r\|_2^{\nu} = \sqrt{\sum_{i=1}^{n} \nu_i ((\Phi r)(i))^2}, \quad \nu_i\text{'s are positive weights.}$$

Proposition 13.

Let Assumptions 1 and 2 hold, and let Π be the projection w.r.t. the weighted Euclidean norm $\|\cdot\|_2^{\xi}$ where ξ is the steady-state probability vector of the Markov chain corresponding to the given policy μ. Then,

(a) The mappings T_μ and ΠT_μ corresponding to μ are contractions of modulus α w.r.t. $\|\cdot\|_2^{\xi}$.

(b) We have

$$\|J_\mu - \Phi r^*\|_2^{\xi} \le \frac{1}{\sqrt{1-\alpha^2}} \|J_\mu - \Pi J_\mu\|_2^{\xi}$$

where J_μ is the unique fixed point of T_μ and r^* is the unique solution of the projected equation $\Phi r = \Pi T_\mu(\Phi r)$.

Proof) Write T_μ in the form $T_\mu J = g_\mu + \alpha P_\mu J$ where g_μ is the vector with elements $\sum_{j=1}^{n} p_{ij}(\mu(i))\, g(i, \mu(i), j)$, $i = 1, \cdots, n$, and P_μ is the matrix with elements $p_{ij}(\mu(i))$.

For all $J, \bar{J} \in \mathbb{R}^n$,

$$\|T_\mu J - T_\mu \bar{J}\|_2^{\xi} = \alpha \|P_\mu (J - \bar{J})\|_2^{\xi} \le \alpha \|J - \bar{J}\|_2^{\xi}.$$

(Why? For $\forall z \in \mathbb{R}^n$, we have

$$\|P_\mu z\|_2^{\xi\,2} = \sum_{i=1}^{n} \xi_i \Big(\sum_{j=1}^{n} p_{ij}(\mu(i)) z_j\Big)^2 \le \sum_{i=1}^{n} \xi_i \sum_{j=1}^{n} p_{ij}(\mu(i)) z_j^2$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} \xi_i p_{ij}(\mu(i)) z_j^2 = \sum_{j=1}^{n} \xi_j z_j^2 = \|z\|_2^{\xi\,2}$$)

Hence, T_μ is a contradiction of modulus α w.r.t. $\|\cdot\|_2^{\xi}$.

For all $J, \bar{J} \in \mathbb{R}^n$,

$$\|\Pi T_\mu J - \Pi T_\mu \bar{J}\|_2^{\xi\,2} = \|\Pi(T_\mu J - T_\mu \bar{J})\|_2^{\xi\,2} \quad (\because \text{ linearity of } \Pi)$$

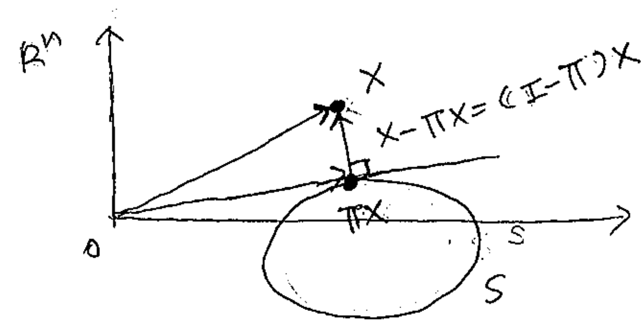
$$\le \|\Pi(T_\mu J - T_\mu \bar{J})\|_2^{\xi\,2} + \|(I - \Pi)(T_\mu J - T_\mu \bar{J})\|_2^{\xi\,2}$$

nonexpansive property of projection

$$= \|T_\mu J - T_\mu \bar{J}\|_2^{\xi\,2} \quad (\because \text{ Pythagorean Theorem})$$

$$\le \alpha^2 \|J - \bar{J}\|_2^{\xi\,2} \quad (\because T_\mu \text{ is contraction of } \alpha \text{ w.r.t. } \|\cdot\|_2^{\xi})$$

Hence, ΠT_μ is also a contraction of modulus α w.r.t. $\|\cdot\|_2^{\xi}$.



$$\|X\|_2^{\xi\,2} = \|\Pi X\|_2^{\xi\,2} + \|(I - \Pi)X\|_2^{\xi\,2} \quad \text{— (24)}$$

Since T_μ and ΠT_μ are contractions, there exist unique fixed points $J_\mu \in \mathbb{R}^n$ and $\Phi r \in S$ satisfying $J_\mu = T_\mu J_\mu$ and $\Phi r = \Pi T_\mu(\Phi r)$, respectively. Moreover, r^* is the unique solution of $\Phi r = \Pi T_\mu(\Phi r)$ by Assumption 2.

$$\|J_\mu - \Phi r^*\|_2^{\xi\,2} = \|J_\mu - \Pi J_\mu\|_2^{\xi\,2} + \|\Pi J_\mu - \Phi r^*\|_2^{\xi\,2}$$

$(\because$ Pythagorean theorem (24) with $X = J_\mu - \Phi r^*)$

$$= \|J_\mu - \Pi J_\mu\|_2^{\xi\,2} + \|\Pi T_\mu J_\mu - \Pi T_\mu(\Phi r^*)\|_2^{\xi\,2}$$

$(\because J_\mu = T_\mu T_\mu$ and $\Phi r^* = \Pi T_\mu(\Phi r^*))$

$$\le \|J_\mu - \Pi J_\mu\|_2^{\xi\,2} + \alpha^2 \|J_\mu - \Phi r^*\|_2^{\xi\,2}$$

$(\because \Pi T_\mu$ is contraction of α w.r.t. $\|\cdot\|_2^{\xi})$

Q.E.D.

Reduced Form of Projected Bellman Equation

- Projected Bellman equation,

$$\Phi r = \Pi T_\mu(\Phi r)$$

is a system of linear equations, since Π is linear and also T is linear of the form

$$T_\mu J = g_\mu + \alpha P_\mu J.$$

- While the system involves n equations with s unknowns (the vector r), it can be reduced to an equivalent set of s equations.

- By the definition of projection, the unique solution r^* must be solution of the following quadractic minimization problem

$$r^* = \arg\min_{r \in R^s} \| \Phi r - (g_\mu + \alpha P_\mu \Phi r) \|_\xi^2$$

$$= \arg\min_{r \in R^s} (\Phi r - (g_\mu + \alpha P_\mu \Phi r))' M (\Phi r - (g_\mu + \alpha P_\mu \Phi r))$$

where $M = \begin{bmatrix} \xi_1 & & \\ & \ddots & \\ & & \xi_n \end{bmatrix}.$

- By setting to 0 the gradient w.r.t. r of the above quadractic objective function, we obtain $\Phi' M (\Phi r - (g + \alpha P_\mu \Phi r)) = 0$.

- Reduced Form of the Projected Bellman Equation

$$\boxed{\Phi' M (\Phi r - (g + \alpha P_\mu \Phi r)) = 0} \qquad (25)$$

This is a system of s linear equations with s unknowns, meaning that $n \times 1$ vector $\Phi r - (g + \alpha P_\mu \Phi r)$ is orthogonal to the columns of Φ in the scaled geometry of the norm $\| \cdot \|_\xi^2$.

- Rewrite the reduced form of the projected Bellman equation (25) in the form of $Ax = b$ as

$$\boxed{\begin{aligned} &\overset{s\times s}{C}\,\overset{s\times 1}{r} = \overset{s\times 1}{d} \\ &\text{where} \quad C = \overset{s\times n}{\Phi'}\,\overset{n\times n}{M}\,(\overset{n\times n}{I - \alpha P_\mu})\,\overset{n\times s}{\Phi}, \quad d = \overset{s\times n}{\Phi'}\,\overset{n\times n}{M}\,\overset{n\times 1}{g_\mu}. \end{aligned}} \quad \text{—— (26)}$$

Note that C is invertible, for otherwise eq. (26) would not have a unique solution, contradicting proposition 13. Thus, eq. (36) can be solved by matrix inversion

$$r = C^{-1} d.$$

$$\left(\begin{aligned} \text{c.f.} \quad \text{Bellman eq.} \quad & J_\mu = T_\mu J_\mu = g_\mu + \alpha P_\mu J_\mu \\ \Rightarrow \quad & \underset{n\times n}{(I - \alpha P_\mu)}\, \underset{n\times 1}{J_\mu} = \underset{n\times 1}{g_\mu} \quad (\text{in the form of } Ax = b) \\ \Rightarrow \quad & J_\mu = (I - \alpha P_\mu)^{-1} g_\mu \end{aligned} \right)$$

- While the reduced form of the projected eq. (26) has smaller dimension than the original Bellman eq. (s rather than n), it may be still hard to solve exactly for large n since computing C and d requires inner product of size n. $\Rightarrow$ Will estimate C and d by using simulation and low-dimensional calculations, e.g., Monte Carlo method.

- Computing C^{-1} for the projected Bellman eq. and $(I - \alpha P_\mu)^{-1}$ for the Bellman eq. is also prohibited in practice for large n and s.

  $\Rightarrow$ Instead, we will consider iterative methods to obtain the solution asymptotically.

$+ \searrow \qquad \Downarrow$

"simulation-based iterative methods"