文章编号:2095-6134(2015)03-0289-06

基于平均化的截尾随机逼近算法*

刘仁龙[1], 杨建奎[1], 熊世峰[2†]

(1 北京邮电大学理学院,北京 100876; 2 中国科学院数学与系统科学研究院, 北京 100190)

(2014 年 4 月 16 日收稿; 2014 年 8 月 6 日收修改稿)

Liu R L, Yang J K, Xiong S F. Averaging-based truncated stochastic approximation algorithm[J]. Journal of University of Chinese Academy of Sciences, 2015,32(3):289-294.

摘 要 考察带有随机干扰线性系统的随机逼近问题. 基于 Polyak 和 Juditsky(SIAM J. Control & Optimization, 1992, 30:838-855)中的平均化加速算法,提出平均化的截尾算法. 证明该算法下随机逼近序列的强相合性和渐近正态性.

关键词 渐近正态性; 线性系统; 强相合性

中图分类号:O226 **文献标志码**:A **doi**:10.7523/j.issn.2095-6134.2015.03.001

Averaging-based truncated stochastic approximation algorithm

LIU Renlong[1], YANG Jiankui[1], XIONG Shifeng[2]

(1 *School of Science*, *Beijing University of Posts and Telecommunications*, *Beijng* 100876, *China*;
2 *Academy of Mathematics and Systems Science*, *Chinese Academy of Sciences*, *Beijing* 100190, *China*)

Abstract In this work the stochastic approximation problem of perturbed linear systems was examined. Inspired by the averaging-based accelerated algorithm of Polyak and Juditsky(SIAM J. Control & Optimization,1992,30:838-855), we propose an averaging-based truncated algorithm. The almost sure convergence and asymptotic normality of the sequence defined by this algorithm are proved.

Key words asymptotic normality; linear system; strong consistency

随机逼近是一种用逐步逼近的方式估计参数的数理统计方法. 随机逼近算法源于文献[1], 后来有了大量的研究, 并被广泛应用于数理统计、随机优化、自适应控制等各领域[2-8]. 随机逼近理论研究的核心问题是其收敛性. 在事先掌握一定先验信息的条件下, 随机逼近序列可以达到理论上的最优收敛速度. Polyak[9] 和 Ruppert[10] 采用平均化方法, 在不需要先验信息的条件下得到具有最优收敛速度的随机逼近算法. Polyak 和 Juditsky[11] 对该算法进行深入研究, 证明对线性或非线性系统, 平均化的随机逼近序列不仅具有强相合性, 而且可以达到理论上的最优收敛速度. 在实际计算中, 由于初值选取、随机误差和异常值等因素的影响, 随机逼近序列可能极大地偏离真值的可行区域, 甚至有时会超出计算机的处理范围. 在这种情况下, 对序列进行截尾是一种

* 国家自然科学基金(11271355,11101050,11471172)资助

† 通信作者,E-mail:xiong@amss.ac.cn

自然的处理方法[12]. 本文针对线性系统，对文献[11]中的算法进行截尾改进. 得到了对应于文献[11]中的收敛结果.

记号和说明：

x^* —方程 $Ax=b$ 的解；

ξ_t —随机误差序列；

x_t —随机逼近序列；

γ_t —迭代的步长；

$I(A)$ —指集合 A 的示性函数.

1 假设及结论

对于线性方程组 $Ax=b$，设其解为 x^*，其中 $b\in R^N$，$A\in R^{N\times N}$，$x\in R^N$，序列 $(y_t)_{t\geq 1}$ 是可以观测到的，此处 y_t 满足

$$y_t = A x_{t-1} - b + \xi_t, \tag{1}$$

其中，$A x_{t-1}-b$ 可以理解为预测残差，ξ_t 是一个随机变量. 为了获得方程的解 x^* 的估计，给出以下算法：

$$x_{t+1}=\begin{cases} x_t-\gamma_{t+1}y_{t+1}, \| x_t-\gamma_{t+1}y_{t+1}\| \leqslant M_{\sigma_t}, \\ M\dfrac{x_t-\gamma_{t+1}y_{t+1}}{\| x_t-\gamma_{t+1}y_{t+1}\|}, \| x_t-\gamma_{t+1}y_{t+1}\| > M_{\sigma_t}, \end{cases} \tag{2}$$

$$(\bar{x}_t) = \frac{1}{t}\sum_{i=0}^{t-1} x_i,$$

$$\sigma_t = \sum_{i=0}^{t-1} I_{[\|x_i-\gamma_{i+1}y_{i+1}\| > M_{\sigma_i}]}, \sigma_0 = 0.$$

其中，M 为一常数，M_t 是一个单调趋于无穷的正实数序列. (2)式可以改写为

$$x_t = z_t I_{(\|z_t\|\leqslant M_{\sigma_t})} + M\frac{z_t}{\|z_t\|} I_{(\|z_t\|>M_{\sigma_t})}, \tag{3}$$

其中

$$z_t = x_{t-1}-\gamma_t y_t.$$

若要估计出 x^*，需要对 γ_t 与 ξ_t 进行一系列的假设如下：

假设 1.1 $\boldsymbol{A}$ 是一个 Hurwitz 矩阵，i. e. ，Re $\lambda_i(\boldsymbol{A})>0$(此处$\lambda_i(\boldsymbol{A})$是 $\boldsymbol{A}$ 的特征值).

假设 1.2 系数(或称步长) γ_t 满足 $\gamma_t\to 0$ 且

$$\frac{\gamma_{t-1}-\gamma_t}{\gamma_t} = o(\gamma_t), \sum_{t=0}^{\infty}\gamma_t^2 < \infty, \sum_{t=0}^{\infty}\gamma_t = \infty. \tag{4}$$

注记 1.1 上式要求步长随着迭代次数的增加，减速不能过快以免跳过最优点.

假设概率空间为 (Ω, F, F_t, P)，设$\boldsymbol{\xi}_t$ 是一个随机变量，并关于 F_t 适应.

假设 1.3 $\boldsymbol{\xi}_t$ 是一个鞅差过程，并且 $\boldsymbol{\xi}_t$ 相互独立，i. e. ，$E(\boldsymbol{\xi}_t \mid F_{t-1})=0$，$\sup_t E(|\boldsymbol{\xi}_t|^2 \mid F_{t-1})<\infty$(a. s.).

a)下列极限存在：

$$\lim_{t\to\infty} E(\boldsymbol{\xi}_t\boldsymbol{\xi}_t^{\mathrm{T}} \mid F_{t-1}) \overset{P}{=\!=} \boldsymbol{S} > 0.$$

其中 $\boldsymbol{\xi}_t$ 是一个列向量，$\boldsymbol{S}>0$ 表示 S 是对称并且正定的.

b)极限存在假设：

$$\lim_{C\to\infty}\overline{\lim_{t\to\infty}} E(|\boldsymbol{\xi}_t|^2 I(|\boldsymbol{\xi}_t|>C) \mid F_{t-1}) \overset{P}{=\!=} 0.$$

下面引入欧氏空间中 2 个集合之间的距离：$\mathrm{d}(S_1,S_2) = \inf\{\|x-y\| : \forall x\in S_1, y\in S_2\}$.

假设 1.4 a) 存在二次连续可微函数 $v(x)$(不一定非负)，记 $v_x(x)$ 为 $v(x)$ 的梯度函数，$h(.)$为 $R^N\to R^N$ 的 Borel 可测函数，在有界集上有界，$x\in J$ 时，$h(x)=0$，则对于 $R^N\to R$，$\forall \triangle > \delta > 0$，都有：

$$\sup_{\triangle > d(x,J) > \delta} h^{\mathrm{T}}(x) v_x(x) < 0.$$

b) $v(J)$ 等于常数 或者 对于任意的 x，只要 $d(x,J)>0$，都有 $d(v(x),v(J))>0$. 其中 $v(J)=\{y : y=v(x), \forall x\in J\}$.

c) $x^\star$ 为 R^N 中，原点为圆心且半径为 M 的球上某一点，设存在常数 c_0 使 $v(x^\star) < \inf_{\|x\|=c_0} v(x)$，同时 $\|x^\star\| < c_0$，并且满足

$$(\bar{v}(J))^c \cap (v(x^\star), \inf_{\|x\|=c_0} v(x)) \neq \varnothing.$$

假设 1.5 记

$$m(k,T) = \max\{m : \sum_{i=k}^{m}\gamma_i \leqslant T\}.$$

对于算法(2)中的收敛子列 $\{x_{n_k}\}$，

$$\lim_{T\to 0}\limsup_{k\to\infty}\frac{1}{T}\|\sum_{i=n_i}^{m(n_k,T_k)}\gamma_i\xi_{i+1}\| = 0, \forall T_k\in[0,T].$$

注记 1.2 假设 1.4、1.5 借用了文献[12]中定理 3.1.1 的条件，在后文中证明截尾序列的有界性会用到. a)、b)、c)中 $v(x)$是存在的，例如当 $N=2$ 时，$\boldsymbol{x}=[x_1,x_2]^{\mathrm{T}}$，令 $h(x)=-x$，$v(x_1,x_2)=\frac{1}{2}(x_1^2+x_2^2)$ 就满足上述条件，读者可自行验证。

如上假设具有一定的合理性，有兴趣的读者可以参考文献[11-13]. 本文沿用文献[13]中扩展截尾随机逼近算法的思路及其假设，不同之处在于本文主要研究截尾后通项算数平均的算法.

定理 1.1　假设 1.1—1.5 满足,则有

$$\sqrt{t}(\overline{x_t}-x^*)\xrightarrow{D}N(0,V).$$

其中, $V=A^{-1}S(A^{-1})^{\mathrm{T}}$, 具体证明过程请参见第 2 部分定理 1.1 的证明.

注记 1.3　定理 1.1 说明对 x_t 进行截尾之后的算术平均仍保持文献[11]中结论成立.

定理 1.2　假设 1.1—1.5 满足则有:

$$\overline{x_t}-x^*\xrightarrow{\text{a. s.}}0.$$

具体证明请参见第 2 部分定理 1.2 的证明.

2　定理的证明

定理 1.1 的证明

记 $\Delta_t=x_t-x^*$, $\overline{\Delta_t}=\dfrac{\sum_{j=0}^{t-1}\Delta_j}{t}$, 则 $\sqrt{t}(x_t-x^*)=\sqrt{t}\Delta_t$, 那么有如下结论:

引理 2.1

$$\sqrt{t}\,\overline{\Delta_t}=\frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\prod_{i=1}^{j}d_i\Delta_0+\frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k}d_if_j+\frac{1}{\sqrt{t}}\sum_{k=0}^{t-1}f_k.\tag{5}$$

其中

$$a_t=I_{(\|z_t\|\leqslant M_{\sigma_t})},$$
$$b_t=I_{(\|z_t\|>M_{\sigma_t})},$$
$$d_t=(I-\gamma_tA)a_t,$$
$$f_t=\left(M\frac{z_t}{\|z_t\|}-x^*\right)b_t-\gamma_t\xi_ta_t.$$

证明　由(1)、(3)可推出:

$$x_t=(x_{t-1}-\gamma_t(Ax_{t-1}-b+\boldsymbol{\xi}_t))I_{(\|z_t\|\leqslant M_{\sigma_t})}+M\frac{z_t}{\|z_t\|}I_{(\|z_t\|>M_{\sigma_t})},\tag{6}$$

两边同时减去 x^* 可以得到

$$\Delta_t=\Delta_{t-1}d_t-c_t+h_t,\tag{7}$$

其中

$$a_t=I_{(\|z_t\|\leqslant M_{\sigma_t})},$$
$$b_t=I_{(\|z_t\|>M_{\sigma_t})},$$
$$d_t=(I-\gamma_tA)a_t,$$
$$c_t=\gamma_t\boldsymbol{\xi}_ta_t,$$
$$h_t=\left(M\frac{z_t}{\|z_t\|}-x^*\right)b_t,$$
$$f_t=h_t-c_t.$$

式(7)通过一些迭代可以获得如下等式:

$$\sqrt{t}\,\overline{\Delta_t}=\frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\prod_{i=1}^{j}d_i\Delta_0+\frac{1}{\sqrt{t}}\sum_{k=0}^{t-1}\sum_{j=1}^{k-1}\prod_{i=j+1}^{k}d_if_j+\frac{1}{\sqrt{t}}\sum_{k=1}^{t-1}f_k,\tag{8}$$

$$\sqrt{t}\,\overline{\Delta_t}=\frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\prod_{i=1}^{j}d_i\Delta_0+\frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k}d_if_j+\frac{1}{\sqrt{t}}\sum_{k=0}^{t-1}f_k.\tag{9}$$

□

进一步整理式(9):

$$\begin{aligned}\sqrt{t}\,\overline{\Delta_t}=&\frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\prod_{i=1}^{j}d_i\Delta_0+\\&\frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_iA)a_ih_j-\\&\frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k}d_i\gamma_j\xi_ja_j+\\&\frac{1}{\sqrt{t}}\sum_{k=0}^{t-1}h_k-\frac{1}{\sqrt{t}}\sum_{k=0}^{t-1}\gamma_k\xi_ka_k\\=&I^{(1)}+I^{(2)}+I^{(3)}+I^{(4)}+I^{(5)}.\end{aligned}\tag{10}$$

下面考察 $I^{(k)}(k=1,2,3,4,5)$.

第 1 步　对于 $I^{(1)}$,

$$\begin{aligned}\|I^{(1)}\|&=\left\|\frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\prod_{i=1}^{j}d_i\Delta_0\right\|\\&=\left\|\frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\prod_{i=1}^{j}(I-\gamma_iA)a_i\Delta_0\right\|\\&\leqslant\left\|\frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\prod_{i=1}^{j}(I-\gamma_iA)\Delta_0\right\|.\end{aligned}$$

定义 $\alpha_j^t=\gamma_j\sum_{k=j}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_iA)$ 且 $\alpha_t=\alpha_0^t$.

则 $\|I^{(1)}\|=\dfrac{1}{\sqrt{t}\gamma_0}\|\alpha_t\Delta_0\|$. 参考文献[11]中引理 2, 在假设 1.2 成立, 且 $\mathrm{Re}\,\lambda_i(A)>0(i=1,2,\cdots,N)$ 的情况下, $\|\alpha_j^t\|$ 有界, 所以有 $|I^{(1)}|=\dfrac{1}{\sqrt{t}\gamma_0}\|\alpha_t\Delta_0\|\to0$ (值收敛).

第 2 步　在考察 $I^{(2)}$ 之前,需要引入引理 2.2.

引理 2.2　在假设 1.4、1.5 满足的条件下,

$$\lim_{n\to\infty}\sigma_n=\sigma<\infty.$$

证明　具体证明可参考文献[12](43 –46 页中关于定理 3.1.1 的证明).

注记 2.1　此引理说明虽然 σ_t 不断增大, 但有界, 这说明在迭代过程中, $(\|z_t\|>M_{\sigma_t})$ 事件发生有限次.

现考察 $I^{(2)}$.

$$
\begin{aligned}
I^{(2)} &= \frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})\,a_i\,h_j \\
&= \frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})\,a_i\Big(M\frac{z_j}{\|z_j\|}-x^*\Big)\,b_j \\
&= \frac{1}{\sqrt{t}}\sum_{j=0}^{t-1}\Big(M\frac{z_j}{\|z_j\|}-x^*\Big)\,b_j\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})\,a_i.
\end{aligned}
\tag{11}
$$

对于($\|z_t\| > M_{\sigma_t}$) 事件发生有限次，不妨设 $b_1,b_2,\cdots,b_L$ 等于 1，$b_i(i=L+1,L+2,\cdots,t-1)$ 等于 0.

$$
\begin{aligned}
\|I^{(2)}\| &= \Big\|\frac{1}{\sqrt{t}}\sum_{j=1}^{L}\Big(M\frac{z_j}{\|z_j\|}-x^*\Big)b_j\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})a_i\Big\| \\
&\leqslant \frac{1}{\sqrt{t}}\sum_{j=1}^{L}\Big\|\Big(M\frac{z_j}{\|z_j\|}-x^*\Big)\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})a_i\Big\| \\
&= \frac{1}{\sqrt{t}}\sum_{j=1}^{L}\Big\|\Big(M\frac{z_j}{\|z_j\|}-x^*\Big)\frac{\alpha_j^t}{\gamma_j}\Big\| \\
&= \frac{1}{\sqrt{t}}\sum_{j=1}^{L}\Big\|\Big(M\frac{z_j}{\|z_j\|}-x^*\Big)\Big\|\,\Big\|\frac{\alpha_j^t}{\gamma_j}\Big\|.
\end{aligned}
\tag{12}
$$

因为 γ_j,α_j^t 有界，且 L 是一常数，所以 $I^{(2)}\to 0$ (值收敛).

第 3 步 考察 $I^{(4)}$

$$
I^{(4)} = \frac{1}{\sqrt{t}}\sum_{k=0}^{t-1}h_k = \frac{1}{\sqrt{t}}\sum_{k=0}^{t-1}\Big(M\frac{z_k}{\|z_k\|}-x^*\Big)\,b_k.
$$

类似于 $I^{(2)}$，$I^{(4)}\to 0$.

第 4 步 考察 $I^{(5)}$

$$
E\,|\,I^{(5)}\,|^2 = E\,\Big|\,\frac{1}{\sqrt{t}}\sum_{k=0}^{t-1}\gamma_k\,\xi_k\,a_k\Big|^2 \leqslant E\,\Big|\,\frac{1}{\sqrt{t}}\sum_{k=1}^{t}\gamma_k\,\xi_k\Big|^2. \tag{13}
$$

根据假设 1.3，$\sup_t E(|\,\xi_t\,|^2\,|\,F_{t-1}) < \infty$(a.s.)，所以存在 $K>0$ 使得 $E\|\,I^{(5)}\,\|^2 \leqslant \frac{K}{t}\sum_{k=1}^{t}\gamma_k^2$，而根据假设 1.2 $\sum_{k=1}^{\infty}\gamma_k^2<\infty$，所以 $\frac{K}{t}\sum_{k=1}^{t}\gamma_k^2\to 0(t\to\infty)$，

所以 $I^{(5)}\to 0$ (均方收敛).

第 5 步 对于 $I^{(3)}$ 需要如下引理:

引理 2.3 记 $P_j^t=\gamma_j\sum_{k=j}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})\,a_i$，$m_j^t=P_j^t-\boldsymbol{A}^{-1}$，存在 $K_1<\infty$, $K_2<\infty$使得 $\frac{1}{t}\sum_{j=0}^{t-1}\|\,m_j^t\,\|\to 0$, $\|\,P_j^t\,\|\leqslant K_1$, $\|\,m_j^t\,\|\leqslant K_2$.

证明 令 $\omega_j^t=\alpha_j^t-\boldsymbol{A}^{-1}$，由($\|z_t\|>M_{\sigma_t}$) 只发生有限次，不妨设 $a_1,a_2,\cdots,a_L$ 为 0，$a_k=1(k=L+1,L+2,\cdots,t-1)$，根据文献[11]中引理 2，$\|\,\alpha_j^t\,\|$ 有界，易知 $\|\,P_j^t\,\|\leqslant\|\,\alpha_j^t\,\|$，所以 $\|\,P_j^t\,\|$ 有界，即存在 $K_1<\infty$使得

$$
\|\,P_j^t\,\|\leqslant K_1.
$$

由 $m_j^t=P_j^t-A^{-1}$，则易知 $\|\,m_j^t\,\|\leqslant\|\,P_j^t\,\|+\|\,A^{-1}\,\|$，所以 $\|\,m_j^t\,\|$ 也有界，故存在 K_2 使得

$$
\|\,m_j^t\,\|\leqslant K_2.
$$

下证 $\sum_{j=0}^{t-1}\|\,m_j^t\,\|\to 0$，有

$$
\begin{aligned}
\sum_{j=0}^{t-1}m_j^t &= \sum_{j=0}^{t-1}(P_j^t-\boldsymbol{A}^{-1}) \\
&= \sum_{j=0}^{t-1}\Big(\gamma_j\sum_{k=j}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})\,a_i-A^{-1}\Big) \\
&= \sum_{j=0}^{L}\Big(\gamma_j\sum_{k=j}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})\,a_i-A^{-1}\Big)+\sum_{j=L+1}^{t-1}\Big(\gamma_j\sum_{k=j}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})\,a_i-A^{-1}\Big) \\
&= -\sum_{j=0}^{L}\boldsymbol{A}^{-1}+\sum_{j=L+1}^{t-1}\Big(\gamma_j\sum_{k=j}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})-\boldsymbol{A}^{-1}\Big).
\end{aligned}
\tag{14}
$$

而

$$
\begin{aligned}
&\sum_{j=L+1}^{t-1}\Big(\gamma_j\sum_{k=j}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})-\boldsymbol{A}^{-1}\Big) \\
&= \sum_{j=1}^{t-1}\Big(\gamma_j\sum_{k=j}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i\boldsymbol{A})-\boldsymbol{A}^{-1}\Big)-\sum_{j=1}^{L}\Big(\gamma_j\sum_{k=j}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_k\boldsymbol{A})-\boldsymbol{A}^{-1}\Big) \\
&= \sum_{j=1}^{t-1}\omega_j^t-\sum_{j=1}^{L}(\alpha_j^t-\boldsymbol{A}^{-1}).
\end{aligned}
\tag{15}
$$

所以有

$$
\frac{1}{t}\sum_{j=0}^{t-1}\|\,m_j^t\,\|\leqslant\frac{1}{t}\sum_{j=1}^{t-1}\|\,\omega_j^t\,\|+\frac{1}{t}\sum_{j=1}^{L}\|\,(\alpha_j^t-\boldsymbol{A}^{-1})\,\|+\frac{1}{t}\sum_{j=0}^{L}\|\,\boldsymbol{A}^{-1}\,\|.
\tag{16}
$$

根据文献[11](846)，$\lim_{t\to\infty}\frac{1}{t}\sum_{j=0}^{t-1}\|\,\omega_j^t\,\|=0$. 而 $\alpha_j^t-\boldsymbol{A}^{-1}$ 有界，所以

$$
\frac{1}{t}\sum_{j=0}^{t-1}\|\,m_j^t\,\|\to 0.
$$

□

对于 $I^{(3)}$，有

$$I^{(3)} = \frac{1}{\sqrt{t}}\sum_{j=1}^{t-1}\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k} d_i\,\gamma_j\,\xi_j\,a_j,$$

因此

$$\begin{aligned}\| I^{(3)} \| &= \| \frac{1}{\sqrt{t}}\sum_{j=1}^{t-1}\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k} d_i\,\gamma_j\,\xi_j\,a_j \| \\ &= \| \frac{1}{\sqrt{t}}\sum_{j=1}^{t-1}\sum_{k=j+1}^{t-1}\prod_{i=j+1}^{k}(I-\gamma_i A)\,a_i\,\gamma_j\,\xi_j\,a_j \| \\ &= \| \frac{1}{\sqrt{t}}\sum_{j=1}^{t-1}(m_j^t + \boldsymbol{A}^{-1})\,\xi_j\,a_j \| \\ &\leqslant \| \frac{1}{\sqrt{t}}\sum_{j=1}^{t-1} m_j^t\,\xi_j\,a_j \| + \| \frac{1}{\sqrt{t}}\sum_{j=1}^{t-1}\boldsymbol{A}^{-1}\,\xi_j\,a_j \|.\end{aligned} \tag{17}$$

对于 $\frac{1}{\sqrt{t}}\sum_{j=1}^{t-1} m_j^t\,\xi_j\,a_j$，

$$E\left| \frac{1}{\sqrt{t}}\sum_{j=1}^{t-1} m_j^t\,\xi_j\,a_j \right|^2$$

$$\leqslant \frac{K}{t}\sum_{j=1}^{t-1}\| m_j^t \|^2 \leqslant \frac{K K_2}{t}\sum_{j=1}^{t-1}\| m_j^t \| \to 0.$$

对于 $\frac{1}{\sqrt{t}}\sum_{j=1}^{t-1} A^{-1}\xi_j a_j$，因为 $(\| z_t \| > M_{\sigma_t})$ 只发生有限次，不妨设 $a_1, a_2, \cdots, a_L$ 为 0，$a_k = 1(k = L+1, L+2, \cdots, t-1)$ 则

$$\frac{1}{\sqrt{t}}\sum_{j=1}^{t-1}\boldsymbol{A}^{-1}\,\xi_j\,a_j = \frac{1}{\sqrt{t}}\sum_{j=1}^{L}\boldsymbol{A}^{-1}\,\xi_j\,a_j + \frac{1}{\sqrt{t}}\sum_{j=L+1}^{t-1}\boldsymbol{A}^{-1}\,\xi_j\,a_j.$$

易得

$$\frac{1}{\sqrt{t}}\sum_{j=1}^{t-1}\boldsymbol{A}^{-1}\,\xi_j\,a_j = \frac{1}{\sqrt{t}}\sum_{j=L+1}^{t-1}\boldsymbol{A}^{-1}\,\xi_j.$$

根据文献[14]中定理 5.5.11，存在一个足够大的常数 C，K_3 有：

$$\overline{\lim_{t\to\infty}}\frac{1}{t}\sum_{j=L+1}^{t-1} E(|\boldsymbol{A}^{-1}\xi_j|^2 I(|\boldsymbol{A}^{-1}\xi_j| > C) \mid F_{j-1}) \leqslant$$

$$K_3^2\,\overline{\lim_{t\to\infty}}\frac{1}{t}\sum_{j=L+1}^{t-1} E(|\xi_j|^2 I(|\xi_j| > C K_3^{-1}) \mid F_{j-1})$$

$$= \varphi(C).$$

根据假设 1.3(b)当 $C\to\infty$时，$\varphi(C)\to 0$，因此满足林德伯格条件，根据中心极限定理 $\frac{1}{\sqrt{t}}\sum_{j=1}^{t-1}\boldsymbol{A}^{-1}\,\xi_j\,a_j$ 依分布收敛到一个正态的随机变量. 因为分解式(10)中其他几部分收敛到零(值收敛或均方收敛)，所以 $\sqrt{t}(\overline{x_t} - x^*)$ 依分布收敛到一个正态的随机变量. 根据假设 1.3(a)，可以得到

$$\frac{1}{t}\sum_{j=1}^{t-1}\boldsymbol{A}^{-1}E(\xi_j\,\xi_j^{\mathrm{T}} \mid F_{j-1})(\boldsymbol{A}^{-1})^{\mathrm{T}}\,a_j^2 \xrightarrow{P} V.$$

综上所述：

$$\sqrt{t}(\overline{x_t} - x^*) \xrightarrow{D} N(0, V).$$

定理 1.2 的证明

证明 利用分解式(10)可以很容易证明 $\frac{I^{(1)}}{\sqrt{t}} \to 0(t\to\infty)$，$\frac{I^{(2)}}{\sqrt{t}} \to 0(t\to\infty)$，$\frac{I^{(4)}}{\sqrt{t}} \to 0(t\to\infty)$，易知 $I^{(3)}$ 的结构和 $I^{(5)}$ 相似，不失一般性只证 $I^{(5)} \to 0(a.s.)$. 对于 $I^{(5)}$ 令 $W_j = \gamma_j\,a_j$.

定义随机序列 $(\xi_t)_{t\geqslant 1}$：

$$\xi_t = \begin{cases}\xi_t, & |\xi_t| \leqslant t^{\frac{3}{4}}, \\ 0, & |\xi_t| > t^{\frac{3}{4}}.\end{cases} \tag{18}$$

根据切比雪夫不等式有：

$$P(|\xi_t| > t^{\frac{3}{4}}) \leqslant E|\xi_t|^2 t^{-\frac{3}{2}} \leqslant K t^{-\frac{3}{2}}. \tag{19}$$

$$\begin{aligned}E\,\xi_j^3 &= E\,\xi_j^2\,\xi_j \\ = \int \xi_j^2\,\xi_j \mathrm{d}F(x) &= \int \xi_j^2\,\xi_j I(|\xi_t| \leqslant t^{\frac{3}{4}})\mathrm{d}F(x) \\ \leqslant t^{\frac{3}{4}}\int \xi_j^2 \mathrm{d}F(x) &= t^{\frac{3}{4}} E\,\xi_j^2 \leqslant K t^{\frac{3}{4}}.\end{aligned} \tag{20}$$

$$\begin{aligned}E\,\xi_j^4 &= E\,\xi_j^2\,\xi_j^2 \\ = \int \xi_j^2\,\xi_j^2 \mathrm{d}F(x) &= \int \xi_j^2\,\xi_j^2 I(|\xi_t| \leqslant t^{\frac{3}{4}})^2\mathrm{d}F(x) \\ \leqslant t^{\frac{3}{2}}\int \xi_j^2 \mathrm{d}F(x) &= t^{\frac{3}{2}} E\,\xi_j^2 \leqslant K t^{\frac{3}{2}}.\end{aligned} \tag{21}$$

下面令 $\frac{1}{t}\sum_{i=1}^{t-1} W_j\,\xi_j = t^{-1} S_t$，易知 W_j 是一致有界的，所以可以充分说明 $t^{-1} S_t \to 0$，

$$\begin{aligned} S_t^4 &= (\sum_{i=1}^{t-1} W_j\,\bar{\xi}_j)^4 = \sum_{i=1}^{t-1}(W_j)^4(\bar{\xi}_j)^4 + \\ & K\sum_{i,j}^{t-1}(W_j)^2\,\bar{\xi}_j^2\,\bar{\xi}_i^2 + \\ & K\sum_{i\neq j, i\neq k, j\leqslant k}^{t-1} W_i\,W_j\,W_k\,\bar{\xi}_i^2\,\bar{\xi}_j\,\bar{\xi}_k + \\ & K\sum_{i<j<k<l}^{t-1} W_i\,W_j\,W_k\,W_l\,\bar{\xi}_i\,\bar{\xi}_j\,\bar{\xi}_k\,\bar{\xi}_l + \\ & K\sum_{i\neq j}^{t-1} W_i(W_j)^3\,\bar{\xi}_i\,\bar{\xi}_j^3 = \sum_{i=1}^{5} I_t^{(i)}.\end{aligned} \tag{22}$$

对于式(22)的第一部分 $I_t^{(1)}$，利用式(21)有

$$E\,I_t^{(1)} = KE\sum_{j=0}^{t-1}\bar{\xi}_j^4 \leqslant K t^{\frac{3}{2}}\sum_{j=0}^{t-1}\bar{\xi}_j^2 \leqslant K t^{\frac{5}{2}}.$$

对于 $I_t^{(5)}$ 利用式(20)有:

$$|EI_t^{(5)}| \leqslant K|E\sum_{i\neq j}^{t-1}\bar{\xi}_i\bar{\xi}_j^3| \leqslant 2K|\sum_{i>j}^{t-1}\bar{\xi}_j^3 E\bar{\xi}_i| \leqslant$$

$$K|\sum_{i>j}^{t-1} j^{\frac{3}{4}} E\bar{\xi}_j^2 i^{-\frac{3}{4}}| \leqslant K|\sum_{i>j}^{t-1}E\bar{\xi}_j^2| \leqslant Kt^2. \tag{23}$$

同理:

$$|EI_t^{(2)}| \leqslant Kt^2, |EI_t^{(3)}| \leqslant Kt^{\frac{3}{2}}, |EI_t^{(4)}| \leqslant Kt.$$

因此

$$t^{-4}ES_t^4 \leqslant Kt^{-4}(t^{\frac{5}{2}} + t^2 + t^{\frac{3}{2}} + t) \leqslant Kt^{-\frac{3}{2}}.$$

根据切比雪夫不等式有

$$\sum_{t=1}^{\infty} P(|t^{-1}S_t| > \delta) \leqslant \sum_{t=1}^{\infty}(t\delta)^{-4}ES_t^4$$

$$\leqslant K\sum_{t=1}^{\infty} t^{-\frac{3}{2}} < \infty.$$

下面证明 $\sum_{t=1}^{\infty} P(|t^{-1}S_t| > \delta) < \infty \Rightarrow \frac{I^{(5)}}{\sqrt{t}} \to 0$ (a. s.).

令事件 $\{\frac{S_t}{t} > \delta\} = A_t$, 则

$$\sum_{t=1}^{\infty} P(A_t) < \infty \Rightarrow \sum_{t=i}^{\infty} P(A_t) \to 0(i\to\infty)$$

$$\Rightarrow P(\bigcup_{t=i}^{\infty} A_t) \to 0(i\to\infty)$$

$$\Rightarrow P(\bigcap_{i=1}^{\infty}\bigcup_{t=i}^{\infty} A_t) \to 0$$

$$\Rightarrow A_t \to 0 \text{ (a. s.)}. \tag{24}$$

所以由大数定理可得 $\frac{I^{(5)}}{\sqrt{t}} \to 0$(a. s.). 所以

$$\bar{x}_t - x^* \to 0\text{(a. s)}.$$

参考文献

[1] Robbins H, Siegmund D. A stochastic approximation method [J]. Ann Math Statist, 1951,22:400-407.

[2] Kiefer E, Wolfovitz J. Stochastic estimation of the maximum of a regression function[J]. Ann Math Statist, 1952,23:462-466.

[3] Kornfelev A P, Stochastic recurrent procedures[M]. Nauka, 1981 (in Russian).

[4] Kushner H J, Clark D S, Stochastic approximation methods for constrained and unconstrained systems[M]. Springer, New York, 1978.

[5] Nemirovski A, Juditsky A, Lan G, et al. Robust stochastic approximation approach to stochastic programming[J]. SIAM J Optim, 2009,19:1 574-1 609.

[6] Stein M L, Chen J, Anitescu M. Stochastic approximation of score functions for Gaussian processes[J]. Ann Appl Stat, 2013,7:1 162-1 191.

[7] Benaim M, Faure M. Stochastic approximation, cooperative dynamics and supermodular games[J]. Ann Appl Probab, 2012,22:2 133-2 164.

[8] Koshal J, Nedic A, Shanbhag U V. Regularized iterative stochastic approximation methods for stochastic variational inequality problems[J]. IEEE Trans Automat Control, 2013, 58:594-609.

[9] Polyak B T. New stochastic approximation type procedures [J]. Avtomat Telemekh, 1990,N7:98-107(in Russian).

[10] Ruppert D. Efficient estimators from a slowly convergent Robbins-Monro process[R]. Tech Report No. 781, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1988.

[11] Polyak B T, Juditsky A B. Acceleration of stochastic approximation by averaging [J]. SIAM J Control and optimization, 1992,30:838-855.

[12] 陈翰馥, 朱允民. 随机逼近[M]. 上海:上海科学技术出版社, 1996.

[13] 陈翰馥, 扩展截尾的随机逼近算法[J]. 系统科学与数学, 2012, 12:1 472-1 487

[14] Liptzer R S, Shiryaev A N. Martingale theory[M]. Nauka, Moscow,1986(in Russian).