Extend [Tsi94] to Eventual Contraction

Longye Tian

July 8, 2025

This note extends Theorem 1 and 3 of [Tsi94] to eventual contraction.

In Section 1, we present a simplified setup of [Tsi94]. In particular, we do not include asynchronous algorithm part. Then we present the proof of Theorem 1 and 3 in [Tsi94]. Moreover, we provide an explicit proof of Lemma 1 in [Tsi94] using Robbins-Siegmund Theorem by [RS71]

In Section 2, we present the extension to eventual contraction by arguing eventual contraction assumption implies original contraction assumption with a specific weighted maximum norm.

Contents

1 Theorem 1 and 3 of [Tsi94]

In this section, we go through a simplified version of Theorem 1 and Theorem 3 of [Tsi94]. In particular, we omit the asynchronous algorithm part. First, we present the simplified setup in [Tsi94][1]. Then, we use Robbins-Siegmund Theorem from [RS71], i.e., Theorem 1.1, to explicitly prove Lemma 1 in [Tsi94] (see Lemma 1.1). Last, we present the proof of Theorem 1 (Theorem 1.2) and 3 (Theorem 1.3) of [Tsi94] in details.

1.1 Simplified model setup and assumptions

Let $x(t)$ denote the state at discrete time $t \in \mathbb{N}$ with component $x_i(t)$. For each component, we have

$$x_i(t+1) = (1 - \alpha_i(t))x_i(t) + \alpha_i(t)(F_i(x(t)) + w_i(t))$$

where

- $\alpha_i(t) \in [0, 1]$ is the stepsize parameter
- $w_i(t)$ is a noise term

All variables are defined on a probability space $(\Omega, \mathcal{F}, P)$ with an increasing sequence of σ-fields $\{\mathcal{F}(t)\}_{t=0}^{\infty}$ representing the algorithm's history.

For any positive vector $v = (v_1, \ldots, v_n)$, we define the weighted maximum norm:

$$\|x\|_v = \max_i \frac{|x_i|}{v_i}, \quad x \in \mathbb{R}^n \tag{1}$$

When $v = (1, \ldots, 1)$, this is the standard maximum norm $\|\cdot\|_\infty$.

Assumption 1.1 (Statistical Properties). We assume

(a) $x(0)$ is $\mathcal{F}(0)$-measurable;

(b) For every i and t, $w_i(t)$ is $\mathcal{F}(t+1)$-measurable;

(c) For every i and t, $\alpha_i(t)$ is $\mathcal{F}(t)$-measurable;

(d) For every i and t, we have $\mathbb{E}[w_i(t) \mid \mathcal{F}(t)] = 0$;

(e) There exist constants A and B such that $\mathbb{E}[w_i^2(t) \mid \mathcal{F}(t)] \leq A + B \max_j \max_{\tau \leq t} |x_j(\tau)|^2$, $\forall i, t$.

Assumption 1.2 (Stepsize conditions). We assume

(a) For every i, $\sum_{t=0}^{\infty} \alpha_i(t) = \infty$, w.p.1;

(b) There exists a constant C such that for every i, $\sum_{t=0}^{\infty} \alpha_i^2(t) \leq C$, w.p.1.

Assumption 1.3 (Contraction). There exists a vector $x^* \in \mathbb{R}^n$, a positive vector v, and a scalar $\beta \in [0, 1)$, such that

$$\|F(x) - x^*\|_v \leq \beta \|x - x^*\|_v, \quad \forall x \in \mathbb{R}^n. \tag{2}$$

Assumption 1.4 (Boundedness). There exists a positive vector v, a scalar $\beta \in [0, 1)$, and a scalar D such that

$$\|F(x)\|_v \leq \beta \|x\|_v + D, \quad \forall x \in \mathbb{R}^n. \tag{3}$$

[1] See Appendix A for direct comparison between the setup in [Tsi94] and the simplified setup

Remark. Notice that Assumption 1.3 implies Assumption 1.4:

$$\begin{aligned}
\|F(x)\|_v &\le \|F(x) - x^*\|_v + \|x^*\|_v && (\Delta \text{ ineq.}) \\
&\le \beta\|x - x^*\|_v + \|x^*\|_v && (\text{Assumption 1.3}) \\
&\le \beta\|x\|_v + (1+\beta)\|x^*\|_v && (\Delta \text{ ineq.})
\end{aligned}$$

Let $D := (1+\beta)\|x^*\|_v$

1.2 Related Theorem

In this section, we present the theorem related to the proof. Currently, we take this theorem as granted. Detailed proof is from [RS71] and here is a very nice blog post related to this theorem, see Why stochastic gradient descent works: The Robbins-Siegmund theorem on almost supermartingales

Theorem 1.1 (Robbins-Siegmund). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be a filtration. Let $\{V_n, \beta_n, \xi_n, \zeta_n\}_{n=0}^{\infty}$ be sequences of non-negative random variables adapted to $\{\mathcal{F}_n\}_{n=0}^{\infty}$ such that:

$$\mathbb{E}[V_{n+1} \mid \mathcal{F}_n] \leq (1 + \beta_n)V_n + \xi_n - \zeta_n \quad \text{a.s. for all } n \geq 0$$

where

- $\sum_{n=0}^{\infty} \beta_n < \infty$ almost surely
- $\sum_{n=0}^{\infty} \xi_n < \infty$ almost surely

Then:

- $\lim_{n\to\infty} V_n = V_\infty$ exists and is finite almost surely
- $\sum_{n=0}^{\infty} \zeta_n < \infty$ almost surely

1.3 Related Lemmas

In the proof of Theorem 1.2, we need to show a noise process under certain conditions converges to zero. This motivates the following lemma. This lemma establishes conditions under which a stochastic process $W(t)$ converges to zero. The process follows the recursion:

$$W(t + 1) = (1 - \alpha(t))W(t) + \alpha(t)w(t)$$

The proof is based on Theorem 1.1.

1. We use the squared process $V(t) = W^2(t)$ and show that the squared process fits the condition of Theorem 1.1.
2. Use Theorem 1.1 to get convergence $V(t) \to V_\infty$
3. Prove $V_\infty = 0$ almost surely by contradiction, hence the original process converges to zero almost surely.

Lemma 1.1. Let $\{\mathcal{F}(t)\}$ be an increasing sequence of σ-fields. For each t, let $\alpha(t)$, $w(t-1)$, and $B(t)$ be $\mathcal{F}(t)$-measurable scalar random variables. Let C be a deterministic constant. Suppose that the following hold with probability 1:

(a) $\mathbb{E}[w(t) \mid \mathcal{F}(t)] = 0$;

(b) $\mathbb{E}[w^2(t) \mid \mathcal{F}(t)] \leq B(t)$;

(c) $\alpha(t) \in [0, 1]$;

(d) $\sum_{t=0}^{\infty} \alpha(t) = \infty$;

(e) $\sum_{t=0}^{\infty} \alpha^2(t) \leq C$.

Suppose that the sequence $\{B(t)\}$ is bounded with probability 1. Let $W(t)$ satisfy the recursion

$$W(t + 1) = (1 - \alpha(t))W(t) + \alpha(t)w(t). \tag{4}$$

Then $\lim_{t\to\infty} W(t) = 0$, with probability 1.

Proof. Let us first note that, without loss of generality, we can assume that $B(t) \leq K$ for some constant K almost surely, since the sequence $\{B(t)\}$ is bounded with probability 1.

Step 1: Use the squared process

We analyze the evolution of the squared process $V(t) = W^2(t)$. From the recursion for $W(t)$, we have:

$$W(t+1) = (1-\alpha(t))W(t) + \alpha(t)w(t)$$

Squaring both sides yields:

$$\begin{aligned} W^2(t+1) &= ((1-\alpha(t))W(t) + \alpha(t)w(t))^2 \\ &= (1-\alpha(t))^2W^2(t) + 2(1-\alpha(t))\alpha(t)W(t)w(t) + \alpha^2(t)w^2(t) \end{aligned}$$

Taking the conditional expectation with respect to $\mathcal{F}(t)$:

$$\mathbb{E}[W^2(t+1) \mid \mathcal{F}(t)] = (1-\alpha(t))^2W^2(t) + 2(1-\alpha(t))\alpha(t)W(t)\mathbb{E}[w(t) \mid \mathcal{F}(t)] + \alpha^2(t)\mathbb{E}[w^2(t) \mid \mathcal{F}(t)]$$

Using the conditions $\mathbb{E}[w(t) \mid \mathcal{F}(t)] = 0$ and $\mathbb{E}[w^2(t) \mid \mathcal{F}(t)] \leq B(t) \leq K$, we obtain:

$$\begin{aligned} \mathbb{E}[V(t+1) \mid \mathcal{F}(t)] &\leq (1-\alpha(t))^2V(t) + \alpha^2(t)K \\ &= (1-2\alpha(t)+\alpha^2(t))V(t) + \alpha^2(t)K \\ &= V(t) - 2\alpha(t)V(t) + \alpha^2(t)V(t) + \alpha^2(t)K \\ &= V(t) - \alpha(t)V(t)(2-\alpha(t)) + \alpha^2(t)K \end{aligned}$$

Since $\alpha(t) \in [0,1]$, we have $(2-\alpha(t)) \geq 1$, which gives:

$$\begin{aligned} \mathbb{E}[V(t+1) \mid \mathcal{F}(t)] &\leq V(t) - \alpha(t)V(t) + \alpha^2(t)K \\ &= (1-\alpha(t))V(t) + \alpha^2(t)K \\ &= V(t) + \alpha^2(t)K - \alpha(t)V(t) \end{aligned}$$

Step 2: Use Theorem 1.1

Now, we let

- $\xi_t = \alpha^2(t)K$, we have
$$\sum_{t=0}^{\infty} \xi_t = \sum_{t=0}^{\infty} \alpha(t)^2 K = K\sum_{t=0}^{\infty} \alpha^2(t) < \infty$$
by our assumption.
- $\zeta_t = \alpha(t)V(t)$ is nonnegative and adapted to the filtration.

Hence, we use Theorem 1.1, we get

- $\lim_{t\to\infty} V(t) = V_\infty$ exists and is finite almost surely
- $\sum_{t=0}^{\infty} \zeta_t = \sum_{t=0}^{\infty} \alpha(t)V(t) < \infty$ almost surely.

Step 3: Prove $V_\infty = 0$ almost surely by contradiction

Suppose that $P(V_\infty \geq 2\epsilon) > \delta$ for some $\epsilon, \delta > 0$. Then we have on the set $\{\omega : V_\infty(\omega) \geq 2\epsilon\}$, by the definition of limit, for every $\omega \in \{\omega : V_\infty(\omega) \geq 2\epsilon\}$, there exists $T(\omega) \in \mathbb{N}$ such that for all $t \geq T(\omega)$, $V(t,\omega) \geq \epsilon$. Hence for all $\omega \in \{V_\infty \geq \epsilon\}$:

$$\sum_{t=0}^{\infty} \zeta_t(\omega) = \sum_{t=0}^{\infty} \alpha(t)V(t,\omega) \geq \sum_{t=T(\omega)}^{\infty} \alpha(t)V(t,\omega) \geq \epsilon \sum_{t=T(\omega)}^{\infty} \alpha(t)$$

By $\sum_{t=0}^{\infty} \alpha(t) = \infty$, we have $\sum_{t=T(\omega)}^{\infty} \alpha(t) = \infty$. Hence

$$\sum_{t=0}^{\infty} \zeta_t(\omega) \geq \epsilon \sum_{t=T(\omega)}^{\infty} \alpha(t) = \infty$$

This implies

$$\left\{\omega : \sum_{t=0}^{\infty} \zeta_t(\omega) = \infty\right\} \supseteq \{\omega : V_\infty(\omega) \geq 2\epsilon\}$$

Hence

$$P\left(\sum_{t=0}^{\infty} \zeta_t = \infty\right) \geq P\left(V_\infty \geq 2\epsilon\right) > \delta$$

This contradicts to $\sum_{t=0}^{\infty} \zeta_t < \infty$ almost surely.

Hence, this contradiction gives $V_\infty = 0$ almost surely. ■

1.4 Main Theorems

This section presents two main theorems from [Tsi94]. The first theorem Theorem 1.2 is to show the stochastic process of interest as discussed in the setup is bounded under assumption 1,2, and 4.

The second main theorem Theorem 1.3 shows that this process converges to zero under assumption 1,2, and 3 which is based on the first theorem.

1.4.1 Theorem 1.2

In this section, we prove the first main theorem Theorem 1.2, which prove the process is bounded. The strategy is proof by contradiction by assuming $x(t)$ is unbounded, in particular,

1. Create a growing envelope $G(t)$ to track the growth of $x(t)$
2. Use this tracking and growing envelope to normalize the noise and this normalized noise fits the condition of Lemma 1.1
3. We use Lemma 1.1 to show that the normalized noise converges to 0
4. Setup the contradiction by selecting a time t_0 that the noise is very small for all $t \geq t_0$
5. Derive the contradiction by showing the growing envelope is stablized after t_0 by induction

Theorem 1.2. Let $(\Omega, \mathcal{F}, P)$ be a probability space with filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$. Let $x(t)$ denote the state at discrete time $t \in \mathbb{N}$ with component $x_i(t)$. For each component, we have

$$x_i(t+1) = (1 - \alpha_i(t))x_i(t) + \alpha_i(t)(F_i(x(t)) + w_i(t))$$

If Assumption 1.1, 1.2, and 1.4 holds, then, the sequence $x(t)$ is bounded with probability 1.

Proof. Step 0: Preliminary setup

First, we assume that we have already discarded a suitable null set, so we do not need to keep repeating the quanlification "with probability 1".

We also assume that all components of the vector v in Assumption 1.4 are equal to 1. (The case of a general positive weighting vector v can be reduced to this special case by a suitable coordinate scaling.)

In other words, we have there exists some $\beta \in [0, 1)$ and some D such that

$$\|F(x)\|_\infty \leq \beta \|x\|_\infty + D, \quad \forall x \in \mathbb{R}^n$$

Step 1: Create a growing envelope $G(t)$ to monitor the growth of $x(t)$

We want to create a growing envelope $G(t)$ to monitor the growth of $x(t)$. Fix $G_0 > 0$ and $\gamma \in [0, 1)$ such that

$$\|F(x)\|_\infty \leq \gamma \max\{\|x\|_\infty, G_0\}, \quad \forall x \in \mathbb{R}^n \tag{5}$$

(Any $\gamma \in [0, 1)$ and $G_0 > 0$ satisfying $\beta G_0 + D \leq \gamma G_0$ will do.)

Then, we fix $\epsilon > 0$ such that $\gamma(1 + \epsilon) = 1$.

> **Remark.** Here $(1 + \epsilon)$ is going to be the growth rate of the envelop. The choice of $\gamma(1 + \epsilon) = 1$ balances out the contraction and growth of the envelope which plays a key role in this proof.

We want this envelope $G(t)$ to bound $x(t)$, we need to find the maximum of $x(t)$ up to some time t, hence, we define

$$M(t) = \max_{\tau \leq t} \|x(\tau)\|_\infty \tag{6}$$

Now, we define a sequence $\{G(t)\}$, recursively.

- Let $G(0) = \max\{M(0), G_0\}$ be the initial bound
- Let

$$G(t+1) = \begin{cases} G(t) & \text{if } M(t+1) \le (1+\epsilon)G(t) \\ G_0(1+\epsilon)^k & \text{if } M(t+1) > (1+\epsilon)G(t) \end{cases} \tag{7}$$

  where k is chosen so that

$$G_0(1+\epsilon)^{k-1} < M(t+1) \le G_0(1+\epsilon)^k$$

Under this construction, we create a growing envelope $G(t)$ such that

- it stays constant unless the process exceeds $(1+\epsilon)$ times of the envelope
- Jumps to the next power of $(1+\epsilon)$ when exceeds this bound.

Under this construction, we also have

$$M(t) \le (1+\epsilon)G(t), \quad \forall t \ge 0 \tag{8}$$

and

$$M(t) \le G(t) \quad \text{if } G(t-1) < G(t) \tag{9}$$

Also notice, that under this contruction, $\{G(t)\}_{t=0}^{\infty}$ is a strictly positive increasing sequence, i.e.,

$$0 < G_0 \le G(0) \le G(1) \le \cdots \tag{10}$$

Step 2: Use this envelope to normalize the noise $w(t)$

Moreover, we have $M(t), G(t)$ are all $\mathcal{F}(t)$-measurable. Next, we define

$$\tilde{w}_i(t) = \frac{w_i(t)}{G(t)}, \quad \forall t \ge 0$$

which is $\mathcal{F}(t+1)$-measurable. Under Assumption 1.1, we have

$$\mathbb{E}(\tilde{w}_i(t)|\mathcal{F}(t)) = \frac{\mathbb{E}(w_i(t)|\mathcal{F}(t))}{G(t)} = 0$$

and

$$\begin{aligned}
\mathbb{E}(\tilde{w}_i^2(t)|\mathcal{F}(t)) &= \frac{\mathbb{E}(w_i^2(t)|\mathcal{F}(t))}{G^2(t)} \\
&\le \frac{A + B\max_j \max_{\tau \le t} |x_j(\tau)|^2}{G^2(t)} && \text{(Assumption 1.1)} \\
&= \frac{A + BM(t)^2}{G^2(t)} && \text{(Equation 6)} \\
&\le \frac{A + B(1+\epsilon)^2 G^2(t)}{G^2(t)} && \text{(Equation 8)} \\
&= \frac{A}{G^2(t)} + B(1+\epsilon)^2 \\
&\le \frac{A}{G_0^2} + B(1+\epsilon)^2 && \text{(Equation 1.4.1)} \\
&=: K \quad \forall t \ge 0
\end{aligned}$$

where K is some deterministic constant.

Step 3: create a recursive structure to use Lemma 1.1

For any i and $t_0 \geq 0$, we define $\tilde{W}_i(t_0; t_0) = 0$ and

$$\tilde{W}_i(t+1; t_0) = (1 - \alpha_i(t))\tilde{W}_i(t; t_0) + \alpha_i(t)\tilde{w}_i(t), \quad \forall t \geq 0 \tag{11}$$

Under this definition, we iterate to get the expression for $\tilde{W}_i(t; 0)$ as

$$\tilde{W}_i(t; 0) = \left[\prod_{\tau=t_0}^{t-1} (1 - \alpha_i(\tau))\right] \tilde{W}_i(t_0; 0) + \tilde{W}_i(t; t_0)$$

for every $t \geq t_0$. This implies

$$|\tilde{W}_i(t; t_0)| \leq |\tilde{W}_i(t; 0)| + |\tilde{W}_i(t_0; 0)|$$

By Lemma 1.1, we have

$$\lim_{t\to\infty} \tilde{W}_i(t; 0) = 0$$

Hence, we have for every $\delta > 0$, there exists some T such that $|\tilde{W}_i(t; t_0)| \leq \delta$, for every t and t_0 satisfying $T \leq t_0 \leq t$.

Step 4: Setup the contradiction

Now we prove that $x(t)$ is bounded by contradiction. Suppose that $x(t)$ is unbounded. The by Equation 6 and Equation 8, we have the tracking envelope $G(t)$ goes to infinity.

Then by the construction of $G(t)$ and Equation 9, the inequality $M(t) \leq G(t)$ holds for infinitely many different values of t.

Moreover, since $G(t)$ goes to infinity and by Equation 9, there exists some t_0 such that for all $t \geq t_0$ the process is bounded by the growing envelope and the noise is very small, i.e., $M(t_0) \leq G(t_0)$ and

$$|\tilde{W}_i(t; t_0)| \leq \epsilon, \quad \forall t \geq t_0, \ \forall i \tag{12}$$

Step 5: Derive the contradiction by showing $G(t)$ stablizes after t_0 via induction

Now we show by induction that for every $t \geq t_0$, we have $G(t) = G(t_0)$ and for every i, we have

$$-G(t_0)(1+\epsilon) \leq -G(t_0) + \tilde{W}_i(t; t_0)G(t_0) \leq x_i(t) \leq G(t_0) + \tilde{W}_i(t; t_0)G(t_0) \leq G(t_0)(1+\epsilon)$$

Base Case $t = t_0$

We start with the case for $t = t_0$, we have

$$|x_i(t_0)| \leq M(t_0) \leq G(t_0)$$

and

$$\tilde{W}_i(t_0; t_0) = 0$$

Hence, we have

$$-G(t_0)(1+\epsilon) \leq -G(t_0) \leq x_i(t) \leq G(t_0) \leq G(t_0)(1+\epsilon)$$

as discussed before, i.e., $x(t)$ is inside the tracking and growing envelope $G(t)$ at $t = t_0$.

Induction hypothesis:

Suppose the result is true for some time $t > t_0$.

Induction Case at $t + 1$

We use this induction hypothesis and Equation 5, we get

$$\begin{aligned}
x_i(t+1) &= (1-\alpha_i(t))x_i(t) + \alpha_i(t)F_i(x(t)) + \alpha_i(t)w_i(t) \\
&\leq (1-\alpha_i(t))(G(t_0) + \tilde{W}_i(t;t_0)G(t_0)) + \alpha_i(t)F_i(x(t)) + \alpha_i(t)w_i(t) && \text{(Induction)} \\
&\leq (1-\alpha_i(t))(G(t_0) + \tilde{W}_i(t;t_0)G(t_0)) + \alpha_i(t)\gamma \max\{\|x\|_\infty, G_0\} + \alpha_i(t)w_i(t) \\
& && \text{(Equation 5)} \\
&\leq (1-\alpha_i(t))(G(t_0) + \tilde{W}_i(t;t_0)G(t_0)) + \alpha_i(t)\gamma G(t_0)(1+\epsilon) + \alpha_i(t)w_i(t) && \text{(Induction)} \\
&\leq (1-\alpha_i(t))(G(t_0) + \tilde{W}_i(t;t_0)G(t_0)) + \alpha_i(t)\gamma G(t_0)(1+\epsilon) + \alpha_i(t)\tilde{w}_i(t)G(t_0) \\
& && \text{(Induction)} \\
&= G(t_0) + ((1-\alpha_i(t))\tilde{W}_i(t;t_0) + \alpha_i(t)\tilde{w}_i(t))G(t_0) && (\gamma(1+\epsilon)=1) \\
&= G(t_0) + \tilde{W}_i(t+1;t_0)G(t_0) && \text{(Equation 11)}
\end{aligned}$$

Symmetrically, we can get the other direction.

Then, by Equation 12, we get

$$|x_i(t+1)| \leq G(t_0)(1+\epsilon)$$

Hence, we have

$$\|x(t+1)\|_\infty \leq G(t_0)(1+\epsilon)$$

By Equation 6, we have $M(t+1) = \max\{M(t), \|x(t+1)\|_\infty\}$. And by the induction hypothesis, we have $G(t) = G(t_0)$ for some $t \geq t_0$. Hence, by Equation 9, we have

$$M(t) \leq (1+\epsilon)G(t) = G(t_0)(1+\epsilon)$$

This implies

$$\begin{aligned}
M(t+1) &= \max\{M(t), \|x(t+1)\|_\infty\} \\
&\leq \max\{G(t_0)(1+\epsilon), G(t_0)(1+\epsilon)\} \\
&= G(t_0)(1+\epsilon) \\
&= G(t)(1+\epsilon)
\end{aligned}$$

Then, by the update rule, i.e., Equation 7, we have $G(t+1) = G(t) = G(t_0)$.

This contradicts to $G(t)$ goes to infinity, hence $x(t)$ is bounded. ■

1.4.2 Theorem 1.3

This is the second main theorem we are interested in, that this process converges to x^* under the contraction assumption, i.e., Assumption 1.3.

The proof of this theorem uses a nested induction approach. The structure of the proof follows

1. Show that $x(t)$ is bounded using Theorem 1.2
2. Create a sequence of decreasing bounds $D_0, D_1, D_2, \cdots$ that converges to zero
3. Prove using induction that for each k, the proess eventually stays within the bounds given by D_k, this is the outer induction.
4. To prove the induction step in the outer induction, we use an inner induction to show that the process eventually moves to D_{k+1}.

Theorem 1.3. Let $(\Omega, \mathcal{F}, P)$ be a probability space with filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$. Let $x(t)$ denote the state at discrete time $t \in \mathbb{N}$ with component $x_i(t)$. For each component, we have

$$x_i(t+1) = (1 - \alpha_i(t))x_i(t) + \alpha_i(t)(F_i(x(t)) + w_i(t))$$

If Assumption 1.1, 1.2, and 1.3 holds,then, the sequence $x(t)$ converges to x^* with probability 1.

Proof. Step 1: Show that $x(t)$ is bounded using Theorem 1.2

Notice that Assumption 1.3 implies Assumption 1.4.

$$\begin{aligned}
\|F(x)\|_v &\le \|F(x) - x^*\|_v + \|x^*\|_v && (\Delta \text{ ineq.})\\
&\le \beta\|x - x^*\|_v + \|x^*\|_v && (\text{Assumption 1.3})\\
&\le \beta\|x\|_v + (1+\beta)\|x^*\|_v && (\Delta \text{ ineq.})
\end{aligned}$$

Let $D := (1+\beta)\|x^*\|_v$. Then we can apply Theorem 1.2, we get $x(t)$ is bounded with probabilty 1. (W.l.o.g., we let $x^* = 0$).

Step 2: Create a sequence of decreasing bounds $D_0, D_1, \cdots$ that converges to 0

Hence there exists some (generally random) D_0 such that

$$\|x(t)\|_\infty \le D_0, \quad \forall t \ge 0$$

Fix some $\epsilon > 0$ such that $\beta(1 + 2\epsilon) < 1$, we define

$$D_{k+1} = \beta(1 + 2\epsilon)D_k, \quad k \ge 0 \tag{13}$$

Clearly, D_k converges to 0.

Step 3: Prove using induction that for each k, the proess eventually stays within the bounds given by D_k

Now we prove by induction to show that for each k, there exists a time t_k such that $\|x(t)\|_\infty \le D_k$ for all $t \ge t_k$.

Base case $k = 0$:
This case has already been shown above.

Induction hypothesis:
Assume there exists t_k such that $\|x(t)\|_\infty \le D_k$ for all $t \ge t_k$.

Induction step at $k + 1$

We need to show there exists $t_{k+1} \geq t_k$ such that

$$\|x(t)\|_\infty \leq D_{k+1}, \quad \forall t \geq t_{k+1}$$

We define a sequence $W_i(t)$ to track the accumulated noise:

- $W_i(0) = 0$
- $W_i(t+1) = (1-\alpha_i(t))W_i(t) + \alpha_i(t)w_i(t)$

As previously shown in the proof of Theorem 1.2, for every $\delta > 0$, there exists a time T such that

$$|W_i(t;t_0)| \leq \delta, \quad \forall T \leq t_0 \leq t \tag{14}$$

Choose $\tau_k \geq t_k$ such that

- $|W_i(t;\tau_k)| \leq \beta\epsilon D_k$ for all $t \geq \tau_k$ by Equation 14

- $\|x(t)\|_\infty \leq D_k$ for all $t \geq \tau_k \geq t_k$ by the induction hypothesis

Then we can define a sequence that provides an upper bound,

- Let $Y_i(\tau_k) = D_k$
- Let
$$Y_i(t+1) = (1-\alpha_i(t))Y_i(t) + \alpha_i(t)\beta D_k \tag{15}$$
for all $t \geq \tau_k$

Step 4: Inner induction

Now we prove that

$$-Y_i(t) + W_i(t;\tau_k) \leq x_i(t) \leq Y_i(t) + W_i(t;\tau_k), \quad \forall t \geq \tau_k \tag{16}$$

by induction.

Inner induction: Base case $t = \tau_k$
This is satisfied by $\|x(t)\|_\infty \leq D_k$ as $Y_i(\tau_k) = D_k$ and $W_i(\tau_k;\tau_k) = 0$.

Inner induction hypothesis:
Suppose that Equation 16 holds for some $t > \tau_k$.

Inner induction step:
Then, we have

$$\begin{aligned}
x_i(t+1) &= (1-\alpha_i(t))x_i(t) + \alpha_i(t)F_i(x(t)) + \alpha_i(t)w_i(t) \\
&\leq (1-\alpha_i(t))(Y_i(t) + W_i(t;\tau_k)) + \alpha_i(t)F_i(x(t)) + \alpha_i(t)w_i(t) && \text{(Inner Induction)} \\
&\leq (1-\alpha_i(t))(Y_i(t) + W_i(t;\tau_k)) + \alpha_i(t)\beta\|x(t)\|_\infty + \alpha_i(t)w_i(t) && \text{(Assumption 1.3)} \\
&\leq (1-\alpha_i(t))(Y_i(t) + W_i(t;\tau_k)) + \alpha_i(t)\beta D_k + \alpha_i(t)w_i(t) && (\|x(t)\|_\infty \leq D_k) \\
&= [(1-\alpha_i(t))Y_i(t) + \alpha_i(t)\beta D_k] + [(1-\alpha_i(t))W_i(t;\tau_k) + \alpha_i(t)w_i(t)] \\
&= Y_i(t+1) + W_i(t+1;\tau_k)
\end{aligned}$$

Symmetrically, we can show the other direction.

Hence, we have

$$-Y_i(t+1) + W_i(t+1) \leq x_i(t+1) \leq Y_i(t+1) + W_i(t+1;\tau_k), \quad \forall t \geq \tau_k$$

Hence, by induction we prove Equation 16.

Finally, we use the result from the inner induction:

Moreover since $Y_i(t)$ are positive, we can get

$$|x_i(t)| \le Y_i(t) + |W_i(t;\tau_k)| \tag{17}$$

By Equation 15, we have

$$\begin{aligned} Y_i(t+1) - \beta D_k &= (1-\alpha_i(t))Y_i(t) + \alpha_i(t)\beta D_k - \beta D_k \\ Z_i(t+1) &= (1-\alpha_i(t))Y_i(t) - (1-\alpha_i(t))\beta D_k && (Z_i(t) := Y_i(t) - \beta D_k) \\ Z_i(t+1) &= (1-\alpha_i(t))Z_i(t) \end{aligned}$$

Hence, by Assumption 1.2, we get $Y_i(t) \to \beta D_k$ as $t \to \infty$.

Since $\limsup$ always exists for bounded sequence, we have

$$\limsup_{t\to\infty} |x_i(t)| \le \limsup_{t\to\infty} (Y_i(t) + |W_i(t;\tau_k)|) \tag{18}$$

Use the property that the limsup of a sum is at most the sum of the limsup, we get

$$\limsup_{t\to\infty} (Y_i(t) + |W_i(t;\tau_k)|) \le \limsup_{t\to\infty} Y_i(t) + \limsup_{t\to\infty} |W_i(t;\tau_k)|$$

As $Y_i(t) \to \beta D_k$, we have $\limsup_{t\to\infty} Y_i(t) = \beta D_k$. Moreover, since we have $|W_i(t;\tau_k)| \le \beta\epsilon D_k$ for all $t \ge \tau_k$ by Equation 14. We have, in all,

$$\limsup_{t\to\infty} |x_i(t)| \le \beta D_k + \beta\epsilon D_k = \beta(1+\epsilon)D_k < \beta(1+2\epsilon)D_k = D_{k+1}$$

by Equation 13. Hence, this implies there exist t large enough, $|x_i(t)| < D_{k+1}$. This proves the claim.

Hence, we have $\|x(t)\|_\infty \le D_k$ for all $t \ge t_k$ and D_k converges to 0. This implies $x(t)$ converges to 0. ■

2 Extension to Eventual Contraction

In this section, we extend the simplified setup in subsection 1.1. In particular, we change the Contraction Assumption 1.3 to Eventual Contraction Assumption 2.3.

The general idea of the proof this extension is

1. show that positive linear operator with spetral radius less than one can be perturbed into a stricly positive linear operator with spetral radius less than one.
2. Use the perturbed strictly positive linear operator, we can define a contraction with weighted maximum norm
3. Use the new contraction to satisfies the original Assumption 1.3 and Assumption 1.4.

2.1 Setup and Assumptions

Let $x(t)$ denote the state at discrete time $t \in \mathbb{N}$ with component $x_i(t)$. For each component, we have

$$x_i(t+1) = (1 - \alpha_i(t))x_i(t) + \alpha_i(t)(F_i(x(t)) + w_i(t))$$

where

- $\alpha_i(t) \in [0, 1]$ is the stepsize parameter
- $w_i(t)$ is a noise term

All variables are defined on a probability space $(\Omega, \mathcal{F}, P)$ with an increasing sequence of σ-fields $\{\mathcal{F}(t)\}_{t=0}^{\infty}$ representing the algorithm's history.

Assumption 2.1 (Statistical Properties). We assume

(a) $x(0)$ is $\mathcal{F}(0)$-measurable;

(b) For every i and t, $w_i(t)$ is $\mathcal{F}(t+1)$-measurable;

(c) For every i and t, $\alpha_i(t)$ is $\mathcal{F}(t)$-measurable;

(d) For every i and t, we have $\mathbb{E}[w_i(t) \mid \mathcal{F}(t)] = 0$;

(e) There exist constants A and B such that $\mathbb{E}[w_i^2(t) \mid \mathcal{F}(t)] \leq A + B \max_j \max_{\tau \leq t} |x_j(\tau)|^2$, $\forall i, t$.

Assumption 2.2 (Stepsize conditions). We assume

(a) For every i, $\sum_{t=0}^{\infty} \alpha_i(t) = \infty$, w.p.1;

(b) There exists a constant C such that for every i, $\sum_{t=0}^{\infty} \alpha_i^2(t) \leq C$, w.p.1.

Assumption 2.3 (Eventual Contraction). There exists a vector $x^* \in \mathbb{R}^n$, and positive linear operator K with spectral radius $\rho(K) < 1$ such that

$$|F(x) - x^*| \leq K|x - x^*|, \quad \forall x \in \mathbb{R}^n. \tag{19}$$

Assumption 2.4 (Boundedness). There exists a positive vector v, a scalar $\beta \in [0, 1)$, and a scalar D such that

$$\|F(x)\|_v \leq \beta \|x\|_v + D, \quad \forall x \in \mathbb{R}^n. \tag{20}$$

2.2 Related Definitions and useful theorems

We use the definition of eventual contraction from DP2. And we use Gelfand's formula to prove that positive linear operator with spectral radius less than one can be perturbed into a strictly positive linear operator with spectral radius less than one as in Lemma 2.2.

Definition 2.1 (Eventual contraction)**.** We call a self-map S on a subset V of a Banach lattice E if there exists a positive linear operator $K : E \to E$ such that

- $\rho(K) < 1$
- $|Sv - Sw| \leq K|v - w|$ for all $v, w \in V$

Lemma 2.1 (Gelfand's formula)**.** If B is any square matrix and $\|\cdot\|$ is any matrix norm, then

$$\rho(B)^k \leq \|B^k\| \quad \text{for all } k \in \mathbb{N}$$

$$\|B^k\|^{1/k} \to \rho(B) \text{ as } k \to \infty$$

Corollary 2.1. If B is any square matrix and $\|\cdot\|$ is any matrix form, then if there exists $n \in \mathbb{N}$ such that

$$\|B^n\| < 1$$

this implies $\rho(B) < 1$.

Proof. Using Lemma 2.1, we have

$$\rho(B)^n \leq \|B^n\| < 1$$

Hence, $\rho(B) < \sqrt[n]{1} = 1$. ■

Lemma 2.2. Let A be a n-dimensional nonnegative square matrix with spectral radius $\rho(A) < 1$. Then there exists a strictly positive matrix B such that

$$A < B \text{ and } \rho(B) < 1$$

Proof. Let J denote the n-dimensional square matrix with every entry equals to 1. We construct $B = A + \epsilon J$. We show that there exists $0 < \epsilon < 1$ such that $\rho(B) < 1$.

Using the Gelfand's formula, we have there exists $N \in \mathbb{N}$ such that for all $n \geq N$, $\|A^n\| < 1$. Fix $n \geq N$. We set $\delta := 1 - \|A^n\|$.

Moreover, we have

$$\begin{aligned}\|B^n\| &= \|(A + \epsilon J)^n\| \\ &= \|A^n + \epsilon(\Gamma_{1,1} + \cdots + \Gamma_{1,C_1^n}) + \cdots + \epsilon^{n-1}(\Gamma_{n-1,1} + \cdots \Gamma_{n-1,C_{n-1}^n}) + \epsilon^n J^n\|\end{aligned}$$

for some square matrix $\Gamma_{i,j}$ and C_j^i be the number of combinations of choosing j objects from i objects.

Remark. To motive this step, we have for $n = 2$,

$$\begin{aligned}(A + \epsilon J)^2 &= A^2 + \epsilon AJ + \epsilon JA + \epsilon^2 J^2 \\ &= A^2 + \epsilon(AJ + JA) + \epsilon^2 J^2\end{aligned}$$

Hence, we have $\Gamma_{1,1} = AJ$ and $\Gamma_{1,2} = JA$ with $C_1^2 = 2$.

Then by triangle inequality, we have

$$\|B^n\| \leq \|A^n\| + \sum_{k=1}^{n-1} \epsilon^k \left(\sum_{j=1}^{C_k^n} \|\Gamma_{k,j}\| \right) + \epsilon^n \|J^n\|$$

Let

$$M := \max_{1 \leq k,j \leq n} \{\|\Gamma_{k,j}\|, \|J^n\|\}$$

$$\gamma := \max_{1 \leq k \leq n} C_k^n$$

By finite dimension, we have M and γ is well-defined and finite. This gives

$$\begin{aligned} \|B^n\| &\leq \|A^n\| + \gamma M \sum_{k=1}^{n} \epsilon^k \\ &< \|A^n\| + \gamma M n \epsilon && (0 < \epsilon < 1) \end{aligned}$$

Let $0 < \epsilon < \frac{\delta}{\gamma M n}$. Then, we have

$$\|B^n\| = \|(A + \epsilon J)^n\| < \|A^n\| + \delta < 1$$

By Corollary 2.1, this implies $\rho(B) < 1$. ∎

2.3 Main extension proof

In this section, we prove a proposition that Assumption 2.3 can be viewed of Assumption 1.3 with a specific weighted maximum norm.

Proposition 2.1 (Assumption 2.3 implies Assumption 1.3)**.** Suppose there exists a vector $x^* \in \mathbb{R}^n$ and a positive linear operator K with spectral radius $\rho(K) < 1$ such that

$$|F(x) - x^*| \leq K|x - x^*|, \quad \forall x \in \mathbb{R}^n$$

Then, this implies there exists a positive vector $v \in \mathbb{R}^n$ and a scalar $\beta \in [0, 1)$, such that

$$\|F(x) - x^*\|_v \leq \beta \|x - x^*\|_v$$

In other words, Assumption 2.3 implies Assumption 1.3.

Proof. First, since K is a positive linear operator in a finite dimensional space, it can be represented by a nonnegative matrix with spectral radius $\rho(K) < 1$.

By Lemma 2.2, there exists a strictly positive matrix $\tilde{K} > K$ such that $\rho(\tilde{K}) < 1$.

Using Perron-Frobenius theorem, we know

- the spectral radius $\beta := \rho(\tilde{K}) = \frac{(\tilde{K}v)_i}{v_i} < 1$ is a positive real simple eigenvalue of $\tilde{K}$
- Its corresponding eigenvector v is uniquely positive up to positive scaling.

Hence, we have pointwise

$$|F_i(x) - x_i^*| \leq (K|x - x^*|)_i \leq (\tilde{K}|x - x^*|)_i, \quad i = 1, 2, \cdots, n$$

as $K < \tilde{K}$. Using the matrix representation, we have

$$(\tilde{K}|x - x^*|)_i = \sum_{j=1}^{n} \tilde{K}_{ij} |x_j - x_j^*|$$

As in Equation 1, we define

$$\|z\|_v := \max_{1 \le i \le n} \frac{|z_i|}{v_i}, \quad \forall z \in \mathbb{R}^n$$

as the weighted maximum norm using v. Hence, this implies

$$\frac{|z_j|}{v_j} \le \max_{1 \le i \le n} \frac{|z_i|}{v_i}, \quad j = 1, 2, \cdots, n$$

Hence,

$$|z_j| \le v_j \|z\|_v, \quad j = 1, 2, \cdots, n$$

We can apply this to $|x_j - x_j^*|$, we get

$$\begin{aligned}
(\tilde{K}|x - x^*|)_i &= \sum_{j=1}^{n} \tilde{K}_{ij} |x_j - x_j^*| \\
&\le \sum_{j=1}^{n} \tilde{K}_{ij} v_j \|x - x^*\|_v \\
&= \|x - x^*\|_v \sum_{j=1}^{n} \tilde{K}_{ij} v_j \\
&= \|x - x^*\|_v (\tilde{K} v)_i
\end{aligned}$$

This implies

$$|F_i(x) - x_i^*| \le \|x - x^*\|_v (\tilde{K} v)_i$$

Now we divide both sides by v_i, we get

$$\frac{|F_i(x) - x_i^*|}{v_i} \le \frac{(\tilde{K} v)_i}{v_i} \|x - x^*\|_v = \beta \|x - x^*\|_v$$

for all $i = 1, 2, \cdots, n$. Hence, we have

$$\|F(x) - x^*\|_v = \max_{1 \le i \le n} \frac{|F_i(x) - x_i^*|}{v_i} \le \beta \|x - x^*\|_v$$

This completes the proof. ■

Remark. Proposition 2.1 implies eventual contraction in finite dimension can be viewed as a contraction with a particular weighted maximum norm. Hence the original Contraction Assumption 1.3 and Boundedness Assumption 1.4 holds under Eventual Contraction Assumption 2.3.

Thus, the original proof for Theorem 1.2 and Theorem 1.3 still holds under eventual contraction.

References

[RS71] Herbert Robbins and David Siegmund. "A convergence theorem for non negative almost supermartingales and some applications". In: *Optimizing methods in statistics*. Elsevier, 1971, pp. 233–257.

[Tsi94] John N Tsitsiklis. "Asynchronous stochastic approximation and Q-learning". In: *Machine learning* 16 (1994), pp. 185–202.

Appendix

A Direct comparison of the simplified setup and [Tsi94]

A.1 Model Setup in [Tsi94]

We consider iterative updates of a vector $x \in \mathbb{R}^n$ to solve the fixed-point equation $F(x) = x$, where $F : \mathbb{R}^n \mapsto \mathbb{R}^n$ with component mappings $F_i : \mathbb{R}^n \mapsto \mathbb{R}$.

Let $x(t)$ denote the state at discrete time $t \in \mathbb{N}$, with components $x_i(t)$. For each component i, we have:

$$x_i(t+1) = \begin{cases} x_i(t), & t \notin T^i \\ x_i(t) + \alpha_i(t)(F_i(x^i(t)) - x_i(t) + w_i(t)), & t \in T^i \end{cases} \tag{21}$$

where:

- $T^i \subset \mathbb{N}$ is the set of update times for component i
- $\alpha_i(t) \in [0, 1]$ is the stepsize parameter
- $w_i(t)$ is a noise term
- $x^i(t) = (x_1(\tau_1^i(t)), \ldots, x_n(\tau_n^i(t)))$ contains possibly outdated information with $0 \leq \tau_j^i(t) \leq t$

All variables are defined on a probability space $(\Omega, \mathcal{F}, P)$ with an increasing sequence of σ-fields $\{\mathcal{F}(t)\}_{t=0}^{\infty}$ representing the algorithm's history.

For any positive vector $v = (v_1, \ldots, v_n)$, we define the weighted maximum norm:

$$\|x\|_v = \max_i \frac{|x_i|}{v_i}, \quad x \in \mathbb{R}^n$$

When $v = (1, \ldots, 1)$, this is the standard maximum norm $\|\cdot\|_\infty$.

A.2 Simplified model setup

Let $x(t)$ denote the state at discrete time $t \in \mathbb{N}$ with component $x_i(t)$. For each component, we have

$$x_i(t+1) = (1 - \alpha_i(t))x_i(t) + \alpha_i(t)(F_i(x(t)) + w_i(t)) \tag{22}$$

where

- $\alpha_i(t) \in [0, 1]$ is the stepsize parameter
- $w_i(t)$ is a noise term

Remark. In this simplified setup, we can write Equation A.2 into vector form, i.e.,

$$\mathbf{x}(t+1) = (I - \mathbf{A}(t))\mathbf{x}(t) + \mathbf{A}(t)(\mathbf{F}(\mathbf{x}(t)) + \mathbf{w}(t))$$

where

$$\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix}, \mathbf{A}(t) = \begin{pmatrix} \alpha_1(t) & 0 & \cdots & 0 \\ 0 & \alpha_2(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \alpha_n(t) \end{pmatrix}, \mathbf{F}(\mathbf{x}(t)) = \begin{pmatrix} F_1(\mathbf{x}(t)) \\ \vdots \\ F_n(\mathbf{x}(t)) \end{pmatrix}, \mathbf{w}(t) = \begin{pmatrix} w_1(t) \\ \vdots \\ w_n(t) \end{pmatrix}$$

But for the sake of doing one thing at a time, we keep the notation in [Tsi94] for now.

All variables are defined on a probability space $(\Omega, \mathcal{F}, P)$ with an increasing sequence of σ-fields $\{\mathcal{F}(t)\}_{t=0}^{\infty}$ representing the algorithm's history.

For any positive vector $v = (v_1, \ldots, v_n)$, we define the weighted maximum norm:

$$\|x\|_v = \max_i \frac{|x_i|}{v_i}, \quad x \in \mathbb{R}^n$$

When $v = (1, \ldots, 1)$, this is the standard maximum norm $\|\cdot\|_\infty$.

Compared to Equation 21, no information is outdated, hence, we have $x^i(t) = x(t)$. And we don't consider the update time for different cases. As mentioned in [Tsi94], this is a special case, hene all the theorems work fine under this setup.

A.3 Assumptions in [Tsi94]

Assumption 1 (Total Asynchronism). For any i and j, $\lim_{t\to\infty} \tau_j^i(t) = \infty$, with probability 1.

Assumption 2 (Statistical Properties).

(a) $x(0)$ is $\mathcal{F}(0)$-measurable;

(b) For every i and t, $w_i(t)$ is $\mathcal{F}(t+1)$-measurable;

(c) For every i, j, and t, $\alpha_i(t)$ and $\tau_j^i(t)$ are $\mathcal{F}(t)$-measurable;

(d) For every i and t, we have $\mathbb{E}[w_i(t) \mid \mathcal{F}(t)] = 0$;

(e) There exist constants A and B such that $\mathbb{E}[w_i^2(t) \mid \mathcal{F}(t)] \leq A + B \max_j \max_{\tau \leq t} |x_j(\tau)|^2$, $\forall i, t$.

Assumption 3 (Stepsize Conditions).

(a) For every i, $\sum_{t=0}^{\infty} \alpha_i(t) = \infty$, w.p.1;

(b) There exists a constant C such that for every i, $\sum_{t=0}^{\infty} \alpha_i^2(t) \leq C$, w.p.1.

Assumption 5 (Contraction). There exists a vector $x^* \in \mathbb{R}^n$, a positive vector v, and a scalar $\beta \in [0, 1)$, such that

$$\|F(x) - x^*\|_v \leq \beta \|x - x^*\|_v, \quad \forall x \in \mathbb{R}^n. \tag{23}$$

Assumption 6 (Boundedness). There exists a positive vector v, a scalar $\beta \in [0, 1)$, and a scalar D such that

$$\|F(x)\|_v \leq \beta \|x\|_v + D, \quad \forall x \in \mathbb{R}^n. \tag{24}$$

> **Remark.** We don't present assumption 4 as it is not required for the theorem of interest. Later, in the simplified version, we will re-enumerate the number. Here is the enumeration to match [Tsi94]

A.4 Simplified assumptions

Assumption 1 (Simplified). This assumption is no longer needed since no information is outdated.

Assumption 2 (Statistical Properties).

(a) $x(0)$ is $\mathcal{F}(0)$-measurable;

(b) For every i and t, $w_i(t)$ is $\mathcal{F}(t+1)$-measurable;

(c) For every i and t, $\alpha_i(t)$ is $\mathcal{F}(t)$-measurable;

(d) For every i and t, we have $\mathbb{E}[w_i(t) \mid \mathcal{F}(t)] = 0$;

(e) There exist constants A and B such that $\mathbb{E}[w_i^2(t) \mid \mathcal{F}(t)] \le A + B \max_j \max_{\tau \le t} |x_j(\tau)|^2$, $\forall i, t$.

Assumption 3 (Stepsize Conditions).

(a) For every i, $\sum_{t=0}^{\infty} \alpha_i(t) = \infty$, w.p.1;

(b) There exists a constant C such that for every i, $\sum_{t=0}^{\infty} \alpha_i^2(t) \le C$, w.p.1.

Assumption 5 (Contraction). There exists a vector $x^* \in \mathbb{R}^n$, a positive vector v, and a scalar $\beta \in [0, 1)$, such that

$$\|F(x) - x^*\|_v \le \beta \|x - x^*\|_v, \quad \forall x \in \mathbb{R}^n. \tag{25}$$

Assumption 6 (Boundedness). There exists a positive vector v, a scalar $\beta \in [0, 1)$, and a scalar D such that

$$\|F(x)\|_v \le \beta \|x\|_v + D, \quad \forall x \in \mathbb{R}^n. \tag{26}$$