Lecture 7–8

Spring 2015 - KTH advanced course

I. PERRON-FROBENIUS

Theorem 1 (Perron-Frobenius)**.** *Let A be an $N \times N$ non-negative matrix with eigenvalues, λ_i. If A is irreducible then:*

(a) There exists $\mathbf{w} > 0$ such that $A\mathbf{w} = \rho(A)\mathbf{w}$, where $\rho(A) = \max_i |\lambda_i|$;

(b) The eigenvector $\mathbf{w}$ is unique up to a scalar multiplication.

Remarks: Let $A \in \mathbb{R}^{N \times N}$ be non-negative and irreducible. Let λ_i be the ith eigenvalue of A with eigenvector $\mathbf{w}_i$. Then:

- The largest eigenvalue of a non-negative, irreducible matrix is positive-real, i.e. λ_i's can be arranged as
$$\lambda_N \geq |\lambda_{N-1}| \geq \ldots \geq \lambda_1.$$
- The eigenvector, $\mathbf{w}_N$, is strictly positive and is unique up to a scalar multiplication. In other words, $|(\mathbf{w}_N)_i| > 0, \forall i$, and each $(\mathbf{w}_N)_i$ has the same sign, where $(\mathbf{w}_N)_i$ is the ith element of $\mathbf{w}_N$.

- For non-negative matrices alone, it is true that there exists an eigenvalue that equals the spectral radius; this eigenvalue may not be simple ($\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$), and may be zero ($\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$). In addition, the corresponding eigenvector may also have zero elements, e.g.
$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$
where the eigenvectors have a zero element.
- For non-negative and irreducible matrices, $\lambda_N > |\lambda_{N-1}|$ is *not* necessarily true, e.g.
$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$
The eigenvalues are $\lambda_{1,2} = -0.5 \pm j0.867, \lambda_3 = 1$; note $|\lambda_{1,2}| = 1$. However, the eigenvector corresponding to $\lambda_3 = 1$ only has to be strictly positive, whereas for other eigenvalues with $|\cdot| = 1$, the eigenvectors may not be strictly positive.

- TRUE or FALSE: For non-negative and irreducible matrices, λ_N is always simple; other eigenvalues can only be equal in magnitude?

Lemma 1. *Let $A \in \mathbb{R}^{n \times n}$ be positive with eigenvalues, λ_i, such that $\lambda_N = \rho(A) = 1$. Then,*

$$\lambda_N > |\lambda_i|, \qquad i \neq N.$$

Proof: Since a positive matrix is non-negative and irreducible, the Perron-Frobenius theorem applies and we have

$$A\mathbf{w} = \lambda_N \mathbf{w} = \mathbf{w}, \qquad \mathbf{w} > 0, \tag{1}$$

where $\mathbf{w}$ is unique up to a scalar multiplication. Suppose there exists an eigenvalue, λ, with

$$A\mathbf{z} = \lambda \mathbf{z}, \qquad |\lambda| = 1, \mathbf{z} \neq 0. \tag{2}$$

Then,

$$0 \leq |\mathbf{z}| = |\lambda \mathbf{z}| = |A\mathbf{z}| \leq |A||\mathbf{z}| = A|\mathbf{z}|.$$

This is because $\mathbf{z}$ is not necessarily non-negative. Define

$$\mathbf{y} \triangleq A|\mathbf{z}| - |\mathbf{z}| \qquad \Rightarrow \qquad \mathbf{y} \geq 0.$$

Suppose, on the contrary, that $\mathbf{y} \neq 0$, then $A\mathbf{y} > 0$; we know that $A|\mathbf{z}| > 0$, then there exists a $\varepsilon > 0$ such that

$$A\mathbf{y} > \varepsilon A|\mathbf{z}|.$$

Thus,

$$A\mathbf{y} = A(A|\mathbf{z}| - |\mathbf{z}|) > \varepsilon A|\mathbf{z}|,$$

or,

$$\begin{aligned} \frac{A}{1+\varepsilon} A|\mathbf{z}| &> A|\mathbf{z}|, \\ \Rightarrow \left(\frac{A}{1+\varepsilon} \right)^k A|\mathbf{z}| &> A|\mathbf{z}|, \qquad k = 1, 2, \ldots. \end{aligned}$$

But $\left(\frac{A}{1+\varepsilon}\right)^k \to 0$ as k increases and the above is true for each k. Thus, $0 > A|\mathbf{z}|$, which is a contradiction (since A is strictly positive), and hence $\mathbf{y} = 0$, resulting into

$$A|\mathbf{z}| = |\mathbf{z}|.$$

From the Perron-Frobenius theorem, Eq. (1), $|\mathbf{z}|$ must be a scalar multiple of $\mathbf{w}$, i.e. $|\mathbf{z}| = \beta \mathbf{w}$. From $A\mathbf{z} = \lambda \mathbf{z}$, we get

$$|A\mathbf{z}| = |\lambda \mathbf{z}| = |\lambda||\mathbf{z}| = |\mathbf{z}|.$$

So we have $|A\mathbf{z}| = |\mathbf{z}|$ and $A|\mathbf{z}| = |\mathbf{z}|$, implying $|A\mathbf{z}| = A|\mathbf{z}|$, which is only true if all elements of $\mathbf{z}$ have the same sign, i.e. $\mathbf{z} = \alpha|\mathbf{z}|, \alpha = \pm 1$. We get

$$\mathbf{z} = \alpha |\mathbf{z}| = \alpha \beta \mathbf{w}.$$

Finally, from Eq. (2):

$$\lambda \mathbf{z} = A\mathbf{z} = A(\alpha\beta\mathbf{w}) = 1 \cdot (\alpha\beta\mathbf{w}),$$

i.e. $\lambda = 1 = \lambda_N$ and the lemma follows. ■

Theorem 2. *Let $A \in \mathbb{R}^{N \times N}$ be primitive (non-negative and irreducible is implied) with eigenvalues, λ_i, then*

$$\lambda_N > |\lambda_i|, \qquad i \neq N,$$

where $\lambda_N = \rho(A)$.

Proof: Since A is primitive, there exists some positive integer, q, such $A^q > 0$. Define $B \triangleq \frac{1}{\lambda_N} A$, then B^q is strictly positive and $\lambda_N(B^q) = \lambda_N(B) = 1$. From Lemma 1,

$$1 = \lambda_N(B^q) > |\lambda_i(B^q)| = |\lambda_i^q(B)| = \left|\frac{\lambda_i^q}{\lambda_N^q}\right|, \qquad \forall i \neq N,$$

leading to

$$\lambda_N > |\lambda_i^q|^{\frac{1}{q}},$$

and the theorem follows. ■

Lemma 2. *Let $A \in \mathbb{R}^{n \times n}$ be non-negative, irreducible with eigenvalues λ_i, such that $\lambda_N = \rho(A)$. Then, λ_N is a simple eigenvalue.*

Proof: Consider $cI_N + A$ for some $c > 0$, then $cI_N + A$ is primitive. For primitive matrices, we know that the largest eigenvalue is positive, real, and strictly greater than every other eigenvalue in magnitude. Suppose on the contrary that λ_N is not simple, then $cI_N + A$ has multiple eigenvalues equaling $c + \lambda_N$, which is a contradiction. Hence, λ_N is simple.

(*I believe that there is another way of proving this statement.*) ■

Corollary 1. *Let $A \in \mathbb{R}^{n \times n}$ be non-negative and irreducible with eigenvalues λ_i, such that $\lambda_N = \rho(A)$. Then, there exists a $c \geq 0$, such that*

$$c + \lambda_N > |c + \lambda_i|, \qquad \forall i \neq N.$$

Proof: Consider $cI_N + A$, which is primitive. Then apply the Perron-Frobenius theorem for primitive matrices. ■

Corollary 2. *Let $A \in \mathbb{R}^{n \times n}$ be non-negative and irreducible with eigenvalues λ_i, such that $\lambda_N = \rho(A)$. Then,*

$$c + \lambda_N > |c + \lambda_i|, \qquad \forall i \neq N,$$

and for any $c > 0$.

A. Eigenspace

Let $A \in \mathbb{R}^{n\times n}$. Then any $\mathbf{v}$ that satisfies $A\mathbf{v} = \lambda\mathbf{v}$ is called the *right* eigenvector of A. Similarly, any $\mathbf{w}$ that satisfies $\mathbf{w}^T A = \lambda\mathbf{w}^T$ is called the *left* eigenvector of A. By definition, the left eigenvectors are the right eigenvectors of A^T. This can be seen by

$$\mathbf{w}^T A = \lambda\mathbf{w}^T \Rightarrow A^T\mathbf{w} = \lambda\mathbf{w}.$$

We call the collection of $\{\mathbf{v}, \lambda\}$ as the *eigenspace* of A and the collection of $\{\mathbf{w}, \lambda\}$ as *eigenspace* of A^T.

For a symmetric matrix, $A = A^T$, the left eigenvectors are the same as the right eigenvectors and thus A and A^T have the same eigenspace. When we decompose a matrix as $A = VDV^{-1}$; the matrix V consists of the right eigenvectors of A and the matrix V^{-1} consists of the left eigenvectors of A (as rows of V^{-1}). This can be shown as

$$A = VDV^{-1} \Rightarrow A^T = V^{-T}DV^T \triangleq WDW^{-1}.$$

Since $A^T = WDW^{-1}$, each column of W is the right eigenvector of A^T. Since $W = (V^{-1})^T$, each column of W is a row in V^{-1}.

A *normal* matrix is such that it can be diagonalized by a diagonal matrix and a unitary matrix ($VV^T = I$, V is unitary real). A symmetric matrix is a normal matrix.

Does A and A^T have the same eigenspace ? Not unless A is normal, i.e., $AA^T = A^TA$. As we have shown above, the relationship between the left and right eigenvectors is given by $W = (V^{-1})^T$. If A is normal, then $V^{-1} = V^T$ and $W = V$.

All of the above can be re-written for complex-valued matrices if we replace the transpose with Hermitian (complex conjugate transpose).

B. Stochastic matrices

A *row(column)-stochastic* matrix is such that it is non-negative and its row(column)-sum is 1.

Lemma 3. *The eigenvalues of a row-stochastic matrix lie in the unit circle.*

Proof: Gershgorin's circle theorem. ■

Lemma 4. *The spectral radius of a row-stochastic matrix is* 1.

Proof: Note that 1 is an eigenvalue and by the above lemma no other eigenvalue exceeds 1. ■

Lemma 5. *The eigenvalues of an irreducible row-stochastic matrix follow:* $|\lambda_1| \leq \ldots \leq |\lambda_{N-1}| \leq \lambda_N = 1$. *The right eigenvector,* $\mathbf{v}_N$*, corresponding to* $\lambda_N = 1$ *is a vector of all constants (positive numbers), i.e.,*

$$\mathbf{v}_N = \frac{1}{\sqrt{N}}\left[1\ 1\ \ldots\ 1\right]^T,$$

after normalization. In addition, if W is primitive (that can be made sure by adding a a strictly positive diagonal) then $|\lambda_{N-1}| < \lambda_N = 1$.

Proof: Perron-Frobenius, W with a strictly positive diagonal is primitive. ■

A *doubly-stochastic* matrix is such that it is both row-stochastic and column-stochastic (or A^T is row-stochastic).

C. Average-consensus algorithm

Consider a strongly-connected graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with N nodes. Let each node possess a real number, $x_i(0)$, at the ith node. Each node implements the following algorithm:

$$x_i(k+1) = \sum_{\{i\}\cup\mathcal{N}_i} w_{ij} x_j(k),$$

where $w_{ij} > 0$ for $i = j$ and $(i, j) \in \mathcal{E}$ such that $\sum_i w_{ij} = 1$. The network-level algorithm can be summarized as

$$\mathbf{x}_{k+1} = W\mathbf{x}_k, \tag{3}$$

where $W = \{w_{ij}\}$ is a weight matrix that collects w_{ij}.

Remark 1. *The weight matrix, W, is row-stochastic and irreducible. With $w_{ii} > 0$, $\forall i$, it is further primitive. From PF theorem, the eigenvalues, λ_i, of W are such that $|\lambda_1| \leq \ldots \leq |\lambda_{N-1}| < \lambda_N = 1$. The right eigenvector, $\mathbf{v}_N$, corresponding to $\lambda_N = 1$ is a strictly positive vector of all constants, i.e.,*

$$\mathbf{v}_N = \frac{1}{\sqrt{N}} \underbrace{[1\ 1\ \ldots\ 1]^T}_{\triangleq \mathbf{1}_N}.$$

Let $\mathbf{v}_i$ be the eigenvector corresponding to λ_i then $W = VDV^{-1}$, where $V = [\mathbf{v}_N,\ \ldots,\ \mathbf{v}_1]$ and D is a diagonal matrix with $\lambda_N, \ldots, \lambda_1$ on the main diagonal.

Consider the asymptotic behavior of (3).

$$\begin{aligned}
\mathbf{x}_{k+1} &= W^{k+1}\mathbf{x}_0, \\
&= VD^{k+1}V^{-1}\mathbf{x}_0, \\
&= [\mathbf{v}_N,\ \ldots,\ \mathbf{v}_1] \begin{bmatrix} 1^{k+1} & & & \\ & \lambda_{N-1}^{k+1} & & \\ & & \ddots & \\ & & & \lambda_1^{k+1} \end{bmatrix} \begin{bmatrix} \underline{\mathbf{v}}_N^T \\ \underline{\mathbf{v}}_{N-1}^T \\ \vdots \\ \underline{\mathbf{v}}_1^T \end{bmatrix} \mathbf{x}_0, \\
&= \mathbf{v}_N \underline{\mathbf{v}}_N^T \mathbf{x}_0 + \sum_{i=1}^{N-1} \lambda_i^{k+1} \mathbf{v}_i \underline{\mathbf{v}}_i^T \mathbf{x}_0, \\
\Rightarrow \mathbf{x}_\infty \triangleq \lim_{k\to\infty} \mathbf{x}_{k+1} &= \mathbf{v}_N \underline{\mathbf{v}}_N^T \mathbf{x}_0.
\end{aligned}$$

If, in addition, W is symmetric then $\mathbf{v}_N = \underline{\mathbf{v}}_N$ and

$$\mathbf{x}_\infty \triangleq \lim_{k\rightarrow\infty} \mathbf{x}_{k+1} \quad = \quad \mathbf{v}_N \underline{\mathbf{v}}_N^T \mathbf{x}_0 = \frac{1}{\sqrt{N}} \mathbf{1}_N \frac{1}{\sqrt{N}} \mathbf{1}_N^T \mathbf{x}_0 = \frac{1}{N} \mathbf{1}\mathbf{1}^T \mathbf{x}_0, \tag{4}$$

where it can be verified that $\mathbf{1}^T\mathbf{x}_0/N$ is the average of the initial condition.

Summary:

Agreement: If $\mathcal{G}$ is strongly-connected and the weights are such that: (i) $w_{ij} > 0$ for all $(i,j) \in \mathcal{E}$ and $(i,i), i \in \mathcal{V}$; and (ii) $\sum_i w_{ij} = 1$; then the update in (3) converges to an agreement over all of the nodes in the network.

Average-consensus: If $\mathcal{G}$ is connected and the weights are such that: (i) $w_{ij} > 0$ for all $(i,j) \in \mathcal{E}$ and $(i,i), i \in \mathcal{V}$; (ii) $\sum_i w_{ij} = 1$; and (iii) $w_{ij} = w_{ji}$; then the update in (3) converges to the average of the nodal initial conditions.

II. Contraction Maps

The following is borrowed from [1]: Chapter 3.

Definition 1 (Contraction)**.** *Consider an iterative algorithm:*

$$\mathbf{x}_{t+1} = T(\mathbf{x}_t), \qquad t \geq 0,$$

where $T : \mathcal{X} \mapsto \mathcal{X}, \mathcal{X} \subseteq \mathbb{R}^n$*, and has the property*

$$\|T(\mathbf{x}) - T(\mathbf{y})\| \leq \alpha \|\mathbf{x} - \mathbf{y}\|, \qquad \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}. \tag{5}$$

Here $\|\cdot\|$ *is some norm, and* $\alpha \in [0, 1)$ *is called the modulus of* T*. Such a mapping is called a contraction mapping and the iterative algorithm is called a contracting iteration.*

Definition 2 (Fixed-point)**.** *Let* $T : \mathcal{X} \mapsto \mathcal{X}$ *be a given mapping. Any vector* $\mathbf{x}^* \in \mathcal{X}$ *satisfying* $T(\mathbf{x}^*) = \mathbf{x}^*$ *is called a fixed-point of* T*.*

Definition 3 (Pseudo-contraction)**.** *Assume that a mapping,* $T : \mathcal{X} \mapsto \mathcal{X}$*, has a fixed point,* $\mathbf{x}^*$*, and is such that*

$$\|T(\mathbf{x}) - \mathbf{x}^*\| \leq \alpha \|\mathbf{x} - \mathbf{x}^*\|, \qquad \forall \mathbf{x} \in \mathcal{X}.$$

Such a mapping, T*, is called a* pseudo-contraction *and the corresponding iterations,* $\mathbf{x}_{t+1} = T(\mathbf{x}_t)$*, are called* pseudo-contracting iterations.

Remark 2.

(i) A mapping, $T : \mathcal{X} \mapsto \mathcal{Y}$*, where* $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^n$ *that satisfies Eq.* (5) *is also called a contraction mapping, even if* $\mathcal{X} \neq \mathcal{Y}$*.*

(ii) If T *is a contraction, then it may not be a pseudo-contraction. Because* T *may not have a fixed-point, e.g. consider* $T : (0, 1] \mapsto (0, 1]$*, defined by* $T(x) = \frac{1}{2}x$*.*

(iii) If T *is a pseudo-contraction, then it may not be a contraction. See the next example.*

(iv) If T *is a contraction and it has a fixed-point, then* T *is a pseudo-contraction.*

(v) Being a contraction is a stronger condition than being a pseudo-contraction.

(vi) Contraction mappings are continuous. In fact, every contraction is Lipschitz continuous but the reverse may not be true, why? Every Lipschitz continuous function is uniformly continuous.

(vii) Pseudo-contractions are not necessarily continuous.

(viii) The iterations $\mathbf{x}_{t+1} = T(\mathbf{x}_t)$ *can be viewed as an algorithm to find a fixed-point. This is because if* $\mathbf{x}_t$ *converges to some* $\mathbf{x}^* \in \mathcal{X}$ *and* T *is continuous at* $\mathbf{x}^*$*, then* $\mathbf{x}^*$ *is a fixed-point of* T*.*

Example 1. *As an example consider, $T : \mathbb{R} \mapsto \mathbb{R}$, defined by $T(x) = \dfrac{x}{2}$. We have*

$$T(x) - T(y) = \frac{1}{2}(x - y), \qquad x, y \in \mathbb{R}.$$

Clearly, T is a contraction with modulus, $\frac{1}{2}$. The iterative algorithm, $x_{t+1} = T(x_t)$, converges to 0, which is a fixed-point of T.

Now consider, $T : [0, 2] \mapsto [0, 2]$, defined by $T(x) = \max\{0, x - 1\}$. Is this a contraction? We have $|T(2) - T(1)| = 1 = |2 - 1|$, and this T is not a contraction. This mapping has a unique *fixed-point, $x^* = 0$, and T is a pseudo-contraction:*

$$T(x) - x^* = \max\{0, x - 1\} \leq \frac{x}{2} = \frac{1}{2}(x - x^*), \qquad \forall x \in [0, 2].$$

See Fig. 3.1.1, taken from [1].

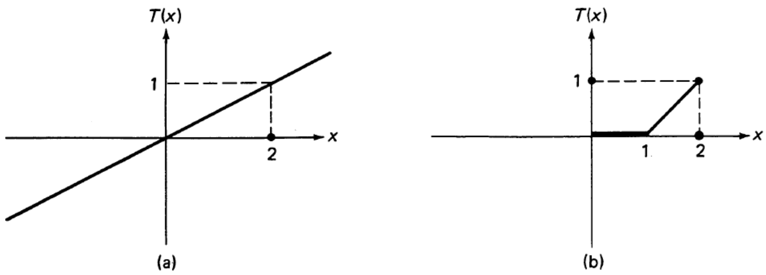

Figure 3.1.1 Illustration of a contraction and a pseudocontraction. (a) The mapping $T : \Re \mapsto \Re$ defined by $T(x) = x/2$ is a contraction with modulus $1/2$, and the iteration $x := T(x)$ converges to zero, which is a fixed point of T. (b) The mapping $T : [0, 2] \mapsto [0, 2]$ defined by $T(x) = \max\{0, x - 1\}$ is not a contraction since $|T(2) - T(1)| = 1$. On the other hand, it has a unique fixed point, equal to zero, and is a pseudocontraction because it is easily seen that $T(x) \leq x/2$ for every $x \in [0, 2]$.

A mapping T can be a contraction for some choice of the vector norm $\|\cdot\|$ and, at the same time, fail to be a contraction under a different choice of norm. Thus, the proper choice of norm is critical.

Lemma 6. *A non-negative matrix, P, has the property, $\rho(P) < 1$, if and only if it is a contraction with respect to some weighted maximum norm.*

This may not be true for non non-negative matrices.

Definition 4. *A Cauchy sequence, $\{\mathbf{x}_t\}_{t \geq 0}$ is such that for every positive real number, $\varepsilon > 0$, there exists a positive integer, N, such that for all integers, $i, j > N$,*

$$\|\mathbf{x}_i - \mathbf{x}_j\| < \varepsilon.$$

Remark 3. *A set is closed if and only if it contains all of its limit points.*

Lemma 7 (Contracting iterations)**.** *Suppose that $T : \mathcal{X} \mapsto \mathcal{X}$ is a contraction with $\alpha \in [0, 1)$ and that $\mathcal{X}$ is a **closed** subset of $\mathbb{R}^n$. Then:*

(i) Existence/uniqueness of fixed-points: The mapping, T, has a unique fixed point, $\mathbf{x}^* \in \mathcal{X}$.

(ii) Geometric convergence: For every initial vector, $\mathbf{x}_0 \in \mathcal{X}$, the sequence, $\{\mathbf{x}_t\}_{t\geq 0}$, generated by $\mathbf{x}_{t+1} = T(\mathbf{x}_t)$ converges to $\mathbf{x}^*$ geometrically. In particular,

$$\|\mathbf{x}_t - \mathbf{x}^*\| \leq \alpha^t \|\mathbf{x}_0 - \mathbf{x}^*\|, \qquad \forall t \geq 0.$$

Proof:

(i) Fix some $\mathbf{x}_0 \in \mathcal{X}$ and consider the sequence $\{\mathbf{x}_t\}$ generated by $\mathbf{x}_{t+1} = T(\mathbf{x}_t)$, then we have $\forall t \geq 0$, (recall $\alpha \in [0, 1)$

$$\|\mathbf{x}_{t+1} - \mathbf{x}_t\| = \|T(\mathbf{x}_t) - T(\mathbf{x}_{t-1})\| \leq \alpha \|\mathbf{x}_t - \mathbf{x}_{t-1}\| \leq \ldots \leq \alpha^t \|\mathbf{x}_1 - \mathbf{x}_0\|.$$

It follows that for every $t \geq 0$ and $m \geq 1$, we have

$$\begin{aligned}
\|\mathbf{x}_{t+m} - \mathbf{x}_t\| &= \left\| \sum_{i=1}^{m} (\mathbf{x}_{t+i} - \mathbf{x}_{t+i-1}) \right\|, \\
&\leq \sum_{i=1}^{m} \|\mathbf{x}_{t+i} - \mathbf{x}_{t+i-1}\|, \\
&\leq \sum_{i=1}^{m} \alpha^{t+i-1} \|\mathbf{x}_1 - \mathbf{x}_0\|, \\
&\leq \alpha^t \sum_{i=0}^{m-1} \alpha^i \|\mathbf{x}_1 - \mathbf{x}_0\|, \\
&\leq \alpha^t \sum_{i=0}^{\infty} \alpha^i \|\mathbf{x}_1 - \mathbf{x}_0\|, \\
&= \frac{\alpha^t}{1-\alpha} \|\mathbf{x}_1 - \mathbf{x}_0\|.
\end{aligned}$$

Therefore, $\{\mathbf{x}_t\}$ is a Cauchy sequence and must converge to a limit, $\mathbf{x}^*$. Furthermore, since $\mathcal{X}$ is closed, the limit of any Cauchy sequence is in $\mathcal{X}$, thus $\mathbf{x}^* \in \mathcal{X}$. We have

$$\|T(\mathbf{x}^*) - \mathbf{x}^*\| \leq \|T(\mathbf{x}^*) - \mathbf{x}_t\| + \|\mathbf{x}_t - \mathbf{x}^*\| \leq \alpha \|\mathbf{x}^* - \mathbf{x}_{t-1}\| + \|\mathbf{x}_t - \mathbf{x}^*\|,$$

for all $t \geq 1$. Since $\mathbf{x}_t$ converges to $\mathbf{x}^*$, we obtain $T(\mathbf{x}^*) = \mathbf{x}^*$. Therefore, the limit, $\mathbf{x}^*$, of $\{\mathbf{x}_t\}$ is a fixed-point of T. It is unique because if $\mathbf{y}^*$ were another fixed-point, we would have

$$\|\mathbf{x}^* - \mathbf{y}^*\| = \|T(\mathbf{x}^*) - T(\mathbf{y}^*)\| \leq \alpha \|\mathbf{x}^* - \mathbf{y}^*\|,$$

which implies that $\mathbf{x}^* = \mathbf{y}^*$.

(ii) We have $\|\mathbf{x}_{t'} - \mathbf{x}^*\| = \|T(\mathbf{x}_{t'-1}) - T(\mathbf{x}^*)\| \leq \alpha \|\mathbf{x}_{t'-1} - \mathbf{x}^*\|$, for all t'; successive application for $t' = t, t-1, \ldots, 1$, obtains the desired result.

■

Lemma 8 (Pseudo-contracting iterations)**.** *Let $\mathcal{X} \subseteq \mathbb{R}^N$. Suppose that $T : \mathcal{X} \mapsto \mathcal{X}$ is a pseudo-contraction with a fixed-point, $\mathbf{x}^* \in \mathcal{X}$, and modulus $\alpha \in [0, 1)$.*

Then, T has no other fixed-points and iterations, $\mathbf{x}_{t+1} = T(\mathbf{x}_t)$ satisfies

$$\|\mathbf{x}_t - \mathbf{x}^*\| \leq \alpha^t \|\mathbf{x}_0 - \mathbf{x}^*\|, \qquad \forall \mathbf{x}_0 \in \mathcal{X}, t \geq 0.$$

In other words, $\mathbf{x}_t$ converges to $\mathbf{x}^*$.

Proof: Uniqueness follows the same argument as Lemma 7, i.e. if $\mathbf{y}^*$ were another fixed-point, then

$$\begin{aligned}
\|\mathbf{x}^* - \mathbf{y}^*\| &= \|T(\mathbf{x}^*) - T(\mathbf{y}^*)\|, \\
&= \|T(\mathbf{x}^*) - \mathbf{y}^* + \mathbf{y}^* - T(\mathbf{y}^*)\|, \\
&\leq \|T(\mathbf{x}^*) - \mathbf{y}^*\| + \|\mathbf{y}^* - T(\mathbf{y}^*)\|, \\
&\leq \alpha \|\mathbf{x}^* - \mathbf{y}^*\|.
\end{aligned}$$

In addition, we have

$$\|\mathbf{x}_t - \mathbf{x}^*\| = \|T(\mathbf{x}_{t-1}) - \mathbf{x}^*\| \leq \alpha \|\mathbf{x}_{t-1} - \mathbf{x}^*\|,$$

for every $t \geq 1$, and the desired result follows by induction on t. ■

In order to apply the above lemma, we need to show that there exists a fixed-point. To this end, Brouwer fixed-point theorem is helpful that comes from purely topological arguments. Another related result is the following:

Lemma 9 (Leray-Schauder-Tychonoff Fixed-Point Theorem)**.** *If $\mathcal{X} \subseteq \mathbb{R}^n$ is non-empty, convex, and compact, and if $T : \mathcal{X} \mapsto \mathcal{X}$ is continuous, then there exists some $\mathbf{x}^* \in \mathcal{X}$ such that $T(\mathbf{x}^*) = \mathbf{x}^*$.*

Remark 4. *If $T : \mathcal{X} \mapsto \mathcal{X}$ is a map with two distinct fixed-points, $\mathbf{x}^*, \mathbf{y}^* \in \mathcal{X}$, then T is not a contraction. This is because*

$$\|T(\mathbf{x}^*) - T(\mathbf{y}^*)\| = \|T(\mathbf{x}^*) - T(\mathbf{y}^*)\|.$$

References

[1] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computations*, Prentice Hall, Englewood Cliffs, NJ, 1989.