# Chapter M

# Monotone convergence

## Contents

## Introduction to monotone convergence theory

Many optimization problems involve finding the minimizer of some **cost function** $\Psi(\boldsymbol{x})$. In such problems, many optimization algorithms monotonically decrease the cost function $\Psi(\boldsymbol{x})$, meaning that $\Psi(\boldsymbol{x}_{n+1}) \leq \Psi(\boldsymbol{x}_n)$ each iteration.

It is of considerable practical importance to determine when this monotonicity property is sufficient for ensuring convergence to a maximizer of $\Psi$, or at least to characterize what additional properties of $\Psi$ are needed to ensure such convergence. Convergence theorems of Ostrowski [7, p. 173] and Meyer [8] are fundamental results in this regard, and are our focus in this section.

Fig. M.1 illustrates why monotonicity alone is not enough to ensure convergence. But these examples are in some sense pathological. Under "reasonable conditions" on $\Psi$ and on an iterative algorithm that monotonically decreases $\Psi$, one can ensure convergence of $\{\boldsymbol{x}_n\}$ to a minimizer.

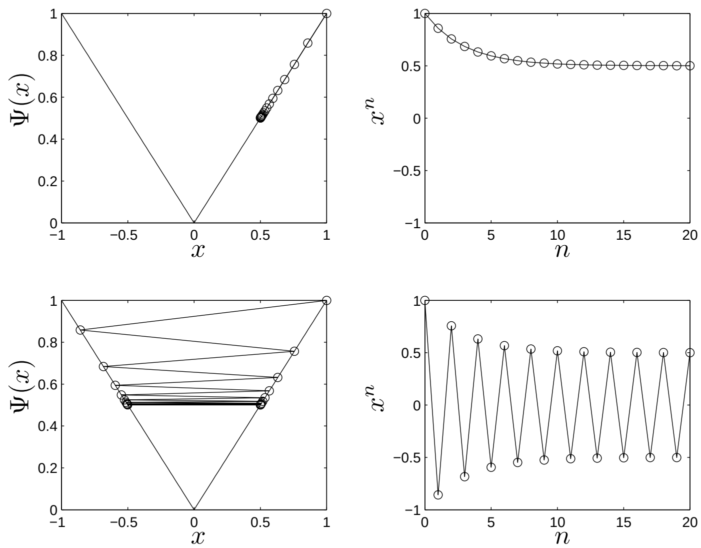


Figure M.1: Monotonicity *alone* need not ensure convergence to a minimizer (top), nor even convergence of $\boldsymbol{x}_n$ (bottom).

## Continuum sets

**Definition.** A closed set *consisting of at least two distinct elements* in a normed space is called a **continuum** iff the set cannot be decomposed into the union of two *nonempty* disjoint closed sets.

(The items in italics are absent from Ostrowski's definition.)

Example. $S = [0, 1]$ is a continuum in $\mathbb{R}$.

Example. $S = [0, 1] \cup [2, 3]$ is not a continuum in $\mathbb{R}$.

## Connected sets

**Definition.** A closed set in a normed space is called **connected** iff it cannot be decomposed into the union of two nonempty disjoint closed sets.

Similarly, an open set in a normed space is called **connected** iff it cannot be decomposed into the union of two nonempty disjoint open sets [9, p. 65]. (There is a more general definition [9, p. 59], but the above is all we need.)

- **Fact.** Any (closed) set that is a continuum is connected.
- A set that is empty or that consists of a single element is connected but is not a continuum.
- A set that is connected but consists of a finite number of points in fact must consist of a single point.

**The set of subsequence limits**

**Definition.** The **set of subsequence limits** of a sequence $\{\boldsymbol{x}_n\}$ in a normed space is the set of limits of convergent subsequences:

$$\boxed{\boldsymbol{x}_n} \triangleq \{\boldsymbol{x} \in \mathcal{X} : \exists \{n_i\} \in \mathbb{N} \text{ (increasing) s.t. } \boldsymbol{x}_{n_i} \to \boldsymbol{x}\}.$$

Example. If $\boldsymbol{x}_n = (1+e^{-n})(-1)^n$, then $\boxed{\boldsymbol{x}_n} = \{-1, 1\}$.

Hereafter, define

$$S_N \triangleq \bigcup_{n=N}^{\infty} \{\boldsymbol{x}_n\}.$$

Then by the definition of $\boxed{\boldsymbol{x}_n}$:

$$\boxed{\boldsymbol{x}_n} \subseteq \overline{S_N}, \qquad \forall N \in \mathbb{N}, \tag{M-1}$$

$$\boldsymbol{x} \in \boxed{\boldsymbol{x}_n} \Longrightarrow d(\boldsymbol{x}, S_N) = 0, \qquad \forall N \in \mathbb{N}. \tag{M-2}$$

As preparation for proving an alternate form of Ostrowski's convergence theorem, we need to show that $\boxed{\boldsymbol{x}_n}$ is a closed set. First we need two lemmas.

This next lemma comes close to being a converse of (M-1).

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

**Lemma 13.7** *If $\{\boldsymbol{x}_n\}$ is a sequence in a normed space, then $d(\boldsymbol{y}, S_N - \{\boldsymbol{y}\}) = 0 \Longrightarrow \boldsymbol{y} \in \boxed{\boldsymbol{x}_n}$ for any $N \in \mathbb{N}$.*

*Proof.* First note that if $M \geq N$ then $0 = d(\boldsymbol{y}, S_N - \{\boldsymbol{y}\}) = \min\{\|\boldsymbol{y} - \boldsymbol{x}_N\|, \ldots, \|\boldsymbol{y} - \boldsymbol{x}_M\|, d(\boldsymbol{y}, S_M - \{\boldsymbol{y}\})\}$.
Thus $d(\boldsymbol{y}, S_M - \{\boldsymbol{y}\}) = 0$ for $M \geq N$ since $\|\boldsymbol{y} - \boldsymbol{x}_n\| > 0$ for $\boldsymbol{y} \neq \boldsymbol{x}_n$.
So we recursively generate $\{n_i\}$ as follows. Pick $n_1 \geq N$ such that $\|\boldsymbol{y} - \boldsymbol{x}_{n_1}\| < 1$.
Having chosen $n_i$, since $d(\boldsymbol{y}, S_{n_i+1}) = 0$, pick $n_{i+1} > n_i$ such that $\left\|\boldsymbol{y} - \boldsymbol{x}_{n_{i+1}}\right\| < 1/(i+1)$.
This $\{n_i\}$ is increasing and $\boldsymbol{x}_{n_i} \to \boldsymbol{y}$ as $i \to \infty$, so $\boldsymbol{y} \in \boxed{\boldsymbol{x}_n}$. □

**Proposition.** In a normed space, any set of subsequence limits $\boxed{\boldsymbol{x}_n}$ is closed.

*Proof.* We show $\widetilde{\boxed{\boldsymbol{x}_n}}$ is open by picking any $\boldsymbol{y} \notin \boxed{\boldsymbol{x}_n}$ and showing that $\boldsymbol{y}$ is an interior point of $\widetilde{\boxed{\boldsymbol{x}_n}}$.

Claim. If $\boldsymbol{y} = \boldsymbol{x}_m$ for some $m \in \mathbb{N}$, then $\boldsymbol{y} = \boldsymbol{x}_n$ for only a finite set of $n$'s.
Pf. If $\boldsymbol{x}_{n_i} = \boldsymbol{y}$ for an infinite set of $n_i$'s, then $\boldsymbol{x}_{n_i} \to \boldsymbol{y}$, contradicting $\boldsymbol{y} \notin \boxed{\boldsymbol{x}_n}$.

Let $M_{\boldsymbol{y}} = 1 + \max\{n \in \mathbb{N} : \boldsymbol{x}_n = \boldsymbol{y}\}$. (Take $M_{\boldsymbol{y}} = 1$ if $\boldsymbol{y} \notin S_1$.)
By the contrapositive of Lemma 13.7, $\boldsymbol{y} \notin \boxed{\boldsymbol{x}_n} \Longrightarrow d\big(\boldsymbol{y}, S_{M_{\boldsymbol{y}}} - \{\boldsymbol{y}\}\big) > 0 \Longrightarrow d\big(\boldsymbol{y}, S_{M_{\boldsymbol{y}}}\big) > 0$ since $\boldsymbol{y} \notin S_{M_{\boldsymbol{y}}}$.
Thus, by Lemma 2.3, $d\big(\boldsymbol{y}, \overline{S_{M_{\boldsymbol{y}}}}\big) > 0$.

So since $\boxed{\boldsymbol{x}_n} \subseteq \overline{S_{M_{\boldsymbol{y}}}}$, as noted in (M-1) above, $d\Big(\boldsymbol{y}, \boxed{\boldsymbol{x}_n}\Big) > 0$, and hence $\boldsymbol{y}$ is an interior point of $\widetilde{\boxed{\boldsymbol{x}_n}}$. □

---

Example. The following sequence in $[0,1] \subset \mathbb{R}$ satisfies the condition $\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| \to 0$.
$\{1/3, 2/3, 3/4, 2/4, 1/4, 1/5, 2/5, 3/5, 4/5, 5/6, 4/6, \ldots\}$.
What is the set of subsequence limits $\boxed{\boldsymbol{x}_n}$ for this sequence? **??**

(Clearly this is not the type of behavior that we would like iterative optimization algorithms to exhibit!)

---

**Ostrowski's convergence theorem: an alternate version**

> **Theorem.** *(Ostrowski, more or less)*
> *Let $\{\boldsymbol{x}_n\}$ be a sequence in a compact subset $K$ of a normed space $(\mathcal{X}, \|\cdot\|)$.*
> *If $\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| \to 0$, then the set of subsequence limits $\boxed{\boldsymbol{x}_n}$ of $\{\boldsymbol{x}_n\}$ is **connected** (and nonempty).*

*Proof.* Let $S_N = \{\boldsymbol{x}_n : n \in \mathbb{N},\ n \geq N\}$. Recall that $\boldsymbol{x} \in \boxed{\boldsymbol{x}_n} \Longrightarrow d(\boldsymbol{x}, S_N) = 0,\ \forall N \in \mathbb{N}$.
$\boxed{\boldsymbol{x}_n}$ is nonempty by the compactness of $K$.

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Claim 0. $\boxed{\boldsymbol{x}_n}$ is compact.
Pf. By construction, $S_1 \subset K$. By Lemma 2.5, $\overline{S_1} \subset K$. By Lemma 2.4, $\overline{S_1}$ is compact.
By (M-1), $\boxed{\boldsymbol{x}_n} \subset \overline{S_1}$. Since $\boxed{\boldsymbol{x}_n}$ is closed by a previous proposition, $\boxed{\boldsymbol{x}_n}$ is compact by Lemma 2.4.

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Suppose $\boxed{\boldsymbol{x}_n} = U \cup V$ where $U$ and $V$ are disjoint nonempty closed sets. We proceed to exhibit a contradiction. ***(Picture)***
Note that $U$ and $V$ are compact since they are closed subsets of $\boxed{\boldsymbol{x}_n}$ and hence closed subsets of $K$, using Lemma 2.4.

By Lemma 2.6, $d(U, V) = \delta > 0$. On the other hand, since $\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| \to 0$, $\exists N_0 \in \mathbb{N}$ s.t. $n > N_0 \Longrightarrow \|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| < \delta/3$.

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Claim 1. For any $N \in \mathbb{N}$, $d(V, S_N) = 0$.
Pf. By (M-2), $d(\boldsymbol{v}, S_N) = 0,\ \forall \boldsymbol{v} \in V$ so $d(V, S_N) = 0,\ \forall N \in \mathbb{N}$.

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Claim 2. $\forall N \in \mathbb{N},\ \exists n > N$ s.t. $d(\boldsymbol{x}_n, V) > 2\delta/3$.
Pf. Suppose $\exists N \in \mathbb{N}$ s.t. $n > N \Longrightarrow d(\boldsymbol{x}_n, V) \leq 2\delta/3$. Pick any $\boldsymbol{u} \in U$ (which is possible since $U$ is nonempty).
Since $d(U, V) = \delta$, we have $d(\boldsymbol{u}, V) \geq \delta$ for $\boldsymbol{u} \in U$.
Since $|d(\boldsymbol{u}, V) - d(\boldsymbol{x}_n, V)| \leq \|\boldsymbol{u} - \boldsymbol{x}_n\|$, we have $\|\boldsymbol{u} - \boldsymbol{x}_n\| \geq d(\boldsymbol{u}, V) - d(\boldsymbol{x}_n, V) \geq \delta - 2\delta/3 = \delta/3$ for all $n > N$.
Thus $d(\boldsymbol{u}, S_{N+1}) \geq \delta/3 > 0$, contradicting (M-2).

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Starting with $n_0 = N_0$, we recursively generate an increasing sequence $\{n_i\}$, as follows, for $i = 1, 2, \ldots$.

By Claim 2, there is some $m > n_{i-1}$ such that $d(\boldsymbol{x}_m, V) > \frac{2}{3}\delta$.
On the other hand, by Claim 1, $\exists n > m$ s.t. $d(\boldsymbol{x}_n, V) < \frac{2}{3}\delta$. Let $n_i$ denote the *smallest* such $n$ so $d(\boldsymbol{x}_{n_i}, V) < \frac{2}{3}\delta$.
Clearly $n_i > m$, so $d(\boldsymbol{x}_{n_i - 1}, V) \geq \frac{2}{3}\delta$.
Furthermore, since $n_i \geq N_0$ in this construction, $d(\boldsymbol{x}_{n_i}, V) \geq d(\boldsymbol{x}_{n_i-1}, V) - \|\boldsymbol{x}_{n_i} - \boldsymbol{x}_{n_i-1}\| > \frac{2}{3}\delta - \frac{1}{3}\delta = \frac{1}{3}\delta$.
Combining, we have $\frac{1}{3}\delta < d(\boldsymbol{x}_{n_i}, V) < \frac{2}{3}\delta$, for an increasing sequence $\{n_i\}$.

Since $\{\boldsymbol{x}_{n_i}\}$ lies in the compact set $K$, it has a convergent subsequence that converges to some limit $\boldsymbol{x} \in K$. Of course $\boldsymbol{x} \in \boxed{\boldsymbol{x}_n}$.
By the continuity of $d(\cdot, V)$, we have $\frac{1}{3}\delta \leq d(\boldsymbol{x}, V) \leq \frac{2}{3}\delta$. Thus $\boldsymbol{x} \notin V$.
On the other hand, by Lemma 2.2, $d(\boldsymbol{x}, U) \geq d(U, V) - d(\boldsymbol{x}, V) > \delta - \frac{2}{3}\delta = \frac{1}{3}\delta$, so $\boldsymbol{x} \notin U$. Thus $\boldsymbol{x} \notin \boxed{\boldsymbol{x}_n}$, which is a contradiction.
Thus, the nonempty closed set $\boxed{\boldsymbol{x}_n}$ cannot be decomposed into the disjoint union of two closed sets, so $\boxed{\boldsymbol{x}_n}$ is connected. □

Example. To see the importance of Ostrowski's conditions, consider the following bizarre sequence in $\mathbb{R}$, for which $\boxed{\boldsymbol{x}_n} = \mathbb{N}$, which is both *not* connected and nonempty (***but what if we relax compactness but require*** $\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| \to 0$***?***):

$$\left(0,\ 1,\ 0 + \tfrac{1}{2},\ 2,\ 0 + \tfrac{3}{4},\ 1 + \tfrac{1}{2},\ 3,\ 0 + \tfrac{7}{8},\ 1 + \tfrac{3}{4},\ 2 + \tfrac{1}{2},\ 4,\ 0 + \tfrac{15}{16},\ 1 + \tfrac{7}{8},\ 2 + \tfrac{3}{4},\ 3 + \tfrac{1}{2},\ \ldots\right).$$

---

**Accumulation points, isolated points, and derived sets**

These concepts are particularly useful for analyzing the convergence of iterative optimization methods.

**Definition.** Let $S$ be a subset of a normed space $(\mathcal{X}, \|\cdot\|)$. A point $\boldsymbol{x} \in \mathcal{X}$ is called an **accumulation point**[1] of $S$ iff

$$\forall \varepsilon > 0,\ \exists \boldsymbol{y} \in S \text{ s.t. } \|\boldsymbol{x} - \boldsymbol{y}\| < \varepsilon \text{ and } \boldsymbol{y} \neq \boldsymbol{x}, \quad \textit{i.e.}, \quad d(\boldsymbol{x}, S - \{\boldsymbol{x}\}) = 0.$$

**Definition.** The set of all accumulation points of a set $S$ is called the **derived set** and is denoted $S'$.

$$S' = \{\boldsymbol{x} \in \mathcal{X} : d(\boldsymbol{x}, S - \{\boldsymbol{x}\}) = 0\}.$$

For comparison, the **closure** of a set is: $\overline{S} = \{\boldsymbol{x} \in \mathcal{X} : d(\boldsymbol{x}, S) = 0\}$. So each **accumulation point** of $S$ is a **cluster point**.

**Definition.** A point $\boldsymbol{x} \in S$ is **isolated** if $d(\boldsymbol{x}, S - \{\boldsymbol{x}\}) > 0$. The collection of isolated points of a set $S$ is:

$$\dot{S} = \{\boldsymbol{x} \in S : d(\boldsymbol{x}, S - \{\boldsymbol{x}\}) > 0\} = \widetilde{S'} \cap S.$$

<u>Example.</u> $S = (0,1) \cup \{2\} \cup [3,4)$
$S' = [0,1] \cup [3,4]$
$\dot{S} = \{2\}$
$\overline{S} = [0,1] \cup \{2\} \cup [3,4]$

<u>Example.</u> $S = \{1, 1/2, 1/3, \ldots\}$
$S' = \{0\}$
$\dot{S} = S$
$\overline{S} = \{0, 1, 1/2, 1/3, \ldots\}$

Simple facts about derived sets.
- $S' \subseteq \overline{S}$
- $\widetilde{S} \cap \overline{S} \subseteq S'$, since $\boldsymbol{x} \notin S \Longrightarrow d(\boldsymbol{x}, S - \{\boldsymbol{x}\}) = d(\boldsymbol{x}, S)$
- $\overline{S} = S \cup S'$
- $S$ is closed iff $S' \subseteq S$
- $\boldsymbol{x} \notin S' \Longrightarrow d(\boldsymbol{x}, S - \{\boldsymbol{x}\}) > 0$.

**Lemma.** $S' = \widetilde{\dot{S}} \cap \overline{S}$, *i.e.*, the derived set is "all of the non-isolated closure points."
*Proof.*

$$\begin{aligned}
\widetilde{\dot{S}} \cap \overline{S} &= \left(S' \cup \widetilde{S}\right) \cap \overline{S} && \text{De Morgan Law} \\
&= (S' \cap \overline{S}) \cup \left(\widetilde{S} \cap \overline{S}\right) && \text{Distributive Law} \\
&= S' \cup \left(\widetilde{S} \cap \overline{S}\right) && \text{since } S' \subseteq \overline{S} \\
&= S' && \text{since } \widetilde{S} \cap \overline{S} \subseteq S'. \quad \square
\end{aligned}$$

.....................................................................................................................

Diagram:

| $\dot{S}$ = isolated points in $S$ | non-isolated points in $S$ | closure points not in $S$ |
|---|---|---|

$S$ is the the first two, $S'$ is the last two, and $\overline{S}$ is all three.
So $S' = \overline{S}$ iff there are no isolated points in $S$.

---

[1] To ensure confusion, a few books call this a **limit point**. We will *not* use that term interchangeably with accumulation points in this course.

**Relation between derived set and subsequence limits**

For a sequence $\{\boldsymbol{x}_n\}$, define $S_N = \cup_{n=N}^{\infty} \{\boldsymbol{x}_n\}$. Then by Lemma 13.7 $S_N' \subseteq \boxed{\boldsymbol{x}_n}, \quad \forall N \in \mathbb{N}$.

Example. Consider the sequence $\boldsymbol{x}_n = (-1)^n$ in $\mathbb{R}$, so $S_N = \{-1, 1\}$.
Then $S_N' = \emptyset$ but $\boxed{\boldsymbol{x}_n} = \{-1, 1\}$.

$S_N'$ and $\boxed{\boldsymbol{x}_n}$ will be the same if the sequence never passes through any of its subsequence limits.

---

**Lemma.** If $S$ is bounded, then $S'$ is bounded.
*Proof.* Since $S$ is bounded, $\exists M < \infty$ s.t. $\|\boldsymbol{x}\| < M$ for all $\boldsymbol{x} \in S$. For any $\boldsymbol{y} \in S'$, $d(\boldsymbol{y}, S - \{\boldsymbol{y}\}) = 0$, so $d(\boldsymbol{y}, S) = 0$.
Thus $\exists \boldsymbol{x} \in S$ s.t. $\|\boldsymbol{y} - \boldsymbol{x}\| < 1$, so $\|\boldsymbol{y}\| \leq \|\boldsymbol{y} - \boldsymbol{x}\| + \|\boldsymbol{x}\| < 1 + M < \infty$. Thus $S'$ is bounded. □

The following proposition is stated to be "obvious" by Ostrowski [7, p. 173].

**Proposition.** In a normed space, any derived set $S'$ is closed.

*Proof.* We show that $\widetilde{S'}$ is open.

Pick any $\boldsymbol{x} \in \widetilde{S'}$. Then $\delta = d(\boldsymbol{x}, S - \{\boldsymbol{x}\}) > 0$.
Now pick a $\boldsymbol{y} \in S - \{\boldsymbol{x}\}$ such that $\delta < \|\boldsymbol{x} - \boldsymbol{y}\| < 2\delta$. (This is possible by definition of the infimum that defines $d$.)
Let $\boldsymbol{z} = \frac{3}{4}\boldsymbol{x} + \frac{1}{4}\boldsymbol{y}$, so $\boldsymbol{z} - \boldsymbol{x} = \frac{1}{4}(\boldsymbol{x} - \boldsymbol{y})$, hence $\delta/4 < \|\boldsymbol{x} - \boldsymbol{z}\| < \delta/2$.

Claim. $B(\boldsymbol{z}, \delta/8) \subset \widetilde{S'}$. (Here we use $B$ to denote an open ball to avoid confusion.)

We need to show that $d(\boldsymbol{t}, S - \{\boldsymbol{t}\}) > 0$ for any $\boldsymbol{t} \in B(\boldsymbol{z}, \delta/8)$.

If $\boldsymbol{t} \in B(\boldsymbol{z}, \delta/8)$ then $\|\boldsymbol{z} - \boldsymbol{t}\| < \delta/8$. Furthermore, $\|\boldsymbol{t} - \boldsymbol{x}\| \leq \|\boldsymbol{t} - \boldsymbol{z}\| + \|\boldsymbol{z} - \boldsymbol{x}\| \leq \delta/8 + \delta/2 = \frac{5}{8}\delta$, and $\|\boldsymbol{t} - \boldsymbol{x}\| \geq \|\boldsymbol{z} - \boldsymbol{x}\| - \|\boldsymbol{t} - \boldsymbol{z}\| \geq \delta/4 - \delta/8 = \delta/8$. Combining: $\delta/8 \leq \|\boldsymbol{t} - \boldsymbol{x}\| \leq \frac{5}{8}\delta$.

$d(\boldsymbol{t}, S - \{\boldsymbol{x}\}) \geq d(\boldsymbol{x}, S - \{\boldsymbol{x}\}) - \|\boldsymbol{t} - \boldsymbol{x}\| \geq \delta - \frac{5}{8}\delta = \frac{3}{8}\delta$. Combining with $d(\boldsymbol{t}, \boldsymbol{x}) \geq \delta/8$, we have $d(\boldsymbol{t}, S) \geq \min\{\delta/8, \frac{3}{8}\delta\} = \delta/8$, so $\boldsymbol{t} \notin S$. Thus, $S = S - \{\boldsymbol{t}\}$ so $d(\boldsymbol{t}, S - \{\boldsymbol{t}\}) \geq \delta/8$, hence $\boldsymbol{t} \in \widetilde{S'}$. Since $\boldsymbol{t}$ within the open ball was arbitrary, we conclude that there is an open ball around $\boldsymbol{z}$ contained in $\widetilde{S'}$, so $\widetilde{S'}$ is open and thus $S'$ is closed. □

***(Picture)*** $[x = 0\ 1/4\ 1/2\ 1\ 2]\delta$ with $y \in [1, 2]$, $z \in [1/4, 1/2]$ and $\boldsymbol{t} \in [1/8, 5/8]$.

***Can you find a simpler proof?***

---

**Lemma.** In a normed space, if $S \subseteq K$ where $K$ is compact, then the following hold.
- $S' \subseteq \overline{S} \subseteq K$
- $\overline{S}$ and $S'$ are also compact.

*Proof.* If $\boldsymbol{x} \in \overline{S}$, then there exists a convergent sequence $\{\boldsymbol{x}_n\} \in S$ such that $\boldsymbol{x}_n \to \boldsymbol{x}$.
Since $\{\boldsymbol{x}_n\} \in K$, there is a convergent subsequence $\{\boldsymbol{x}_{n_i}\}$ such that $\boldsymbol{x}_{n_i} \to \boldsymbol{y}$ for some $\boldsymbol{y} \in K$.
But since $\{\boldsymbol{x}_n\}$ is convergent, its subsequences converge to the same limit, so $\boldsymbol{x} = \boldsymbol{y} \in K$. Since $\boldsymbol{x}$ was arbitrary, $\overline{S} \subseteq K$.
A previous lemma showed $S' \subseteq \overline{S}$, so $S' \subseteq \overline{S} \subseteq K$.

Since both $\overline{S}$ and $S'$ are closed sets that are subsets of the compact set $K$, they are each compact by Lemma 2.4. □

---

Example. The following sequence in $[0, 1] \subset \mathbb{R}$ satisfies the condition of Ostrowski's theorem below:
$\{1/3, 2/3, 3/4, 2/4, 1/4, 1/5, 2/5, 3/5, 4/5, 5/6, 4/6, \ldots\}$. What is the derived set for this sequence? **??**

(Clearly this is not the type of behavior that we would like iterative optimization algorithms to exhibit!)

---

**Ostrowski's convergence theorem**

Often with iterative algorithms we see that as the iterations proceed, the iterates $\boldsymbol{x}_n$ change less and less. Does this mean that $\{\boldsymbol{x}_n\}$ converges? The following convergence theorem of Ostrowski partially addresses this question.

(Originally stated in Euclidean space. We generalize here to arbitrary normed spaces.)

**Theorem.** *(Ostrowski)*
*Let $\{\boldsymbol{x}_n\}$ be a sequence that lies within a compact subset $K$ of a normed space $(\mathcal{X}, \|\cdot\|)$ and let $S = \{\boldsymbol{x}_n : n \in \mathbb{N}\}$.*
*If $\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| \to 0$, then either*
- *$\{\boldsymbol{x}_n\}$ converges in the usual sense (to some limit $\boldsymbol{x}_\star \in K$), or*
- *the derived set $S'$ of $S$ is a continuum.*

*Proof.* Since $\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| \to 0$, if $S$ is finite then it is clear that $\{\boldsymbol{x}_n\}$ must converge in the usual sense.
So we focus on the case that $S$ is infinite hereafter.

$S'$ is closed and compact as shown in a previous proposition and lemma.

Now we suppose $S' = U \cup V$ where $U$ and $V$ are disjoint nonempty closed sets, and we proceed to form a contradiction.
Note that $U$ and $V$ are compact since they are subsets of $S'$ and hence subsets of $K$, using a previous lemma.

By Lemma 2.6, $d(U, V) = \delta > 0$.
On the other hand, since $\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| \to 0$, $\exists N_0 \in \mathbb{N}$ s.t. $n > N_0 \implies \|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| < \delta/3$.

---

Claim 1. For any $N \in \mathbb{N}$ and $\forall \boldsymbol{x} \in S'$, $d(\boldsymbol{x}, S_N - \{\boldsymbol{x}\}) = 0$, where $S_N \triangleq \bigcup_{n=N}^{\infty} \{\boldsymbol{x}_n\}$. (This union is valid since $|S| = \infty$.)
Pf. Suppose $d(\boldsymbol{x}, S_N - \{\boldsymbol{x}\}) = \varepsilon > 0$ for some $\boldsymbol{x} \in S'$. Then $\forall n \geq N,\ \boldsymbol{x}_n \neq \boldsymbol{x} \implies \|\boldsymbol{x}_n - \boldsymbol{x}\| \geq \varepsilon$.
But then $d(\boldsymbol{x}, S - \{\boldsymbol{x}\}) = \min\left\{d(\boldsymbol{x}, S_N - \{\boldsymbol{x}\}),\ d\left(\boldsymbol{x}, \{\boldsymbol{x}_n\}_{n=1}^{N-1} - \{\boldsymbol{x}\}\right)\right\} > 0$, which contradicts the assumption that $\boldsymbol{x} \in S'$.

---

Claim 2. For any $N \in \mathbb{N}$, $d(S_N, V) = 0$.
Pf. By Claim 1, $d(\boldsymbol{v}, S_N) = 0,\ \forall \boldsymbol{v} \in V$ so $d(V, S_N) = 0,\ \forall N \in \mathbb{N}$.

---

Claim 3. $\forall N \in \mathbb{N},\ \exists n > N$ s.t. $d(\boldsymbol{x}_n, V) > 2\delta/3$.
Pf. Suppose $\exists N \in \mathbb{N}$ s.t. $n > N \implies d(\boldsymbol{x}_n, V) \leq 2\delta/3$. Pick any $\boldsymbol{x} \in U$ (which is possible since $U$ is nonempty).
Since $d(U, V) = \delta$, we have $d(\boldsymbol{x}, V) \geq \delta$ for $\boldsymbol{x} \in U$.
Since $|d(\boldsymbol{x}, V) - d(\boldsymbol{x}_n, V)| \leq \|\boldsymbol{x} - \boldsymbol{x}_n\|$, we have $\|\boldsymbol{x} - \boldsymbol{x}_n\| \geq d(\boldsymbol{x}, V) - d(\boldsymbol{x}_n, V) \geq \delta - 2\delta/3 = \delta/3$ for all $n > N$.
Thus $d(\boldsymbol{x}, S_{N+1}) \geq \delta > 0$, contradicting Claim 1.

---

Starting with $n_0 = N_0$, we recursively generate an increasing sequence $\{n_i\}$, as follows, for $i = 1, 2, \ldots$.

By Claim 3, there is some $m > n_{i-1}$ such that $d(\boldsymbol{x}_m, V) > \frac{2}{3}\delta$ and hence $\|\boldsymbol{y} - \boldsymbol{x}_m\| > \frac{2}{3}\delta$.
On the other hand, by Claim 2, $\exists n > m$ s.t. $d(\boldsymbol{x}_n, V) < \frac{2}{3}\delta$. Let $n_i$ denote the *smallest* such $n$ so $d(\boldsymbol{x}_{n_i}, V) < \frac{2}{3}\delta$.
Clearly $n_i > m$, so $d(\boldsymbol{x}_{n_i-1}, V) \geq \frac{2}{3}\delta$.
Furthermore, since $n_i \geq N_0$ in this construction, $d(\boldsymbol{x}_{n_i}, V) \geq d(\boldsymbol{x}_{n_i-1}, V) - \|\boldsymbol{x}_{n_i} - \boldsymbol{x}_{n_i-1}\| \geq \frac{2}{3}\delta - \frac{1}{3}\delta = \frac{1}{3}\delta$.
Combining, we have $\frac{1}{3}\delta \leq d(\boldsymbol{x}_{n_i}, V) < \frac{2}{3}\delta$, for an increasing sequence $\{n_i\}$.

Since $\{\boldsymbol{x}_{n_i}\}$ lies in the compact set $\overline{S}$, it has a convergent subsequence that converges to some limit $\boldsymbol{x} \in \overline{S}$.
By the continuity of $d(\cdot, V)$, we have $\frac{1}{3}\delta \leq d(\boldsymbol{x}, V) < \frac{2}{3}\delta$. Thus $\boldsymbol{x} \notin V$.
On the other hand, $d(\boldsymbol{x}, U) \geq d(U, V) - d(\boldsymbol{x}, V) \geq \delta - \frac{2}{3}\delta = \frac{1}{3}\delta$, so $\boldsymbol{x} \notin U$. Thus $\boldsymbol{x} \notin S'$.

Now, $\boldsymbol{x}_{n_i} \to \boldsymbol{x} \notin S'$, so take the subsequence $\{\boldsymbol{x}_{n_i}\}$ and remove $\boldsymbol{x}$ from it if necessary to form a sub-subsequence, call it $\{\boldsymbol{x}_{m_i}\}$, for which $\boldsymbol{x}_{m_i} \to \boldsymbol{x}$ as $i \to \infty$, but $\boldsymbol{x}_{m_i} \neq \boldsymbol{x},\ \forall i \in \mathbb{N}$. Thus $0 = d(\boldsymbol{x}, \{\boldsymbol{x}_{m_i}\}) = d(\boldsymbol{x}, \{\boldsymbol{x}_{m_i}\} - \{\boldsymbol{x}\}) \geq d(\boldsymbol{x}, S - \{\boldsymbol{x}\}) \geq 0$, so necessarily $\boldsymbol{x} \in S'$. This is a contradiction! Thus, either $\{\boldsymbol{x}_n\}$ converges, or $S'$ must be a continuum. □

***(Picture)*** ?

---

**Meyer's monotone convergence theorem**

Consider the problem of finding a minimizer of a *continuous* cost function $\Psi : C \to \mathbb{R}$, *i.e.*

$$\arg\min_{\boldsymbol{x} \in S} \Psi(\boldsymbol{x}),$$

where $C$ is a *closed* set in a normed space $(\mathcal{X}, \|\cdot\|)$, $S$ is a compact set (so that $\Psi$ achieves its minimum over $S$), and $S \subseteq C$.

A 1976 paper by Meyer [8] provides theorems that give quite general conditions under which an algorithm that monotonically decreases the cost function will converge. Meyer's result is quite general and considers point-to-set mappings. For simplicity, we focus on the common case where the iterative algorithm is defined as follows:

$$\boldsymbol{x}_{n+1} = T(\boldsymbol{x}_n), \tag{M-3}$$

where $T : C \to C$ is a *continuous* operator.

To state Meyer's result, we first make a few definitions.

**Definition.** A point $\boldsymbol{x}_\star \in \mathcal{X}$ is a **fixed point** of $T$ iff $\boldsymbol{x}_\star = T(\boldsymbol{x}_\star)$ . Let $T_\star \triangleq \{\boldsymbol{x} \in \mathcal{X} : T(\boldsymbol{x}) = \boldsymbol{x}\}$ denote the set of fixed points.

**Definition.** An operator $T$ is called **strictly monotone** (decreasing) on $C$ with respect to a cost function $\Psi(\boldsymbol{x})$ iff
- $\Psi(T(\boldsymbol{x})) \leq \Psi(\boldsymbol{x}), \ \forall \boldsymbol{x} \in C$, and
- $\Psi(T(\boldsymbol{x})) < \Psi(\boldsymbol{x})$ whenever $\boldsymbol{x}$ is *not* a fixed point of $T$, *i.e.*, $\forall \boldsymbol{x} \in \widetilde{T_\star} \cap C$.

**Definition.** An operator $T$ is called **uniformly compact** on $C$ iff there exists a compact set $K$ such that $T(\boldsymbol{x}) \in K$ for all $\boldsymbol{x} \in C$.

Example. $T(x) = \sin x$ with $K = [-1, 1]$ is uniformly compact.

The usual case of interest is when $C = \{\boldsymbol{x} \in \mathcal{X} : \Psi(\boldsymbol{x}) \leq \Psi(\boldsymbol{x}_0)\}$ is compact.

**Theorem.** *(Meyer [8])*
*Suppose $T : C \to C$ is a continuous operator over a closed set $C$ in a normed space $(\mathcal{X}, \|\cdot\|)$ such that*

$$T \text{ is uniformly compact on } C\text{, and} \quad \text{(M-4)}$$
$$T \text{ is strictly monotonic on } C\text{, with respect to a continuous function } \Psi. \quad \text{(M-5)}$$

*If $\{\boldsymbol{x}_n\}$ is any sequence generated by the algorithm (M-3) corresponding to $T$ with $\boldsymbol{x}_0 \in C$, then*

1. *all subsequence limit points (of $\{\boldsymbol{x}_n\}$) will be fixed points (of $T$),* i.e., $\boxed{\boldsymbol{x}_n} \subseteq T_\star$ (M-6)
2. $\Psi(\boldsymbol{x}_n) \to \Psi(\boldsymbol{x}_\star)$, *where $\boldsymbol{x}_\star$ is a fixed point of $T$,* i.e., $\boldsymbol{x}_\star \in T_\star$ (M-7)
3. $\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| \to 0$, *and* (M-8)
4. $\boxed{\boldsymbol{x}_n}$*, the set of subsequence limits of $\{\boldsymbol{x}_n\}$, is connected.* (M-9)

*Proof.* Clearly $\{\boldsymbol{x}_n\} \in C$. Since $T$ is uniformly compact, $\exists K$ compact such that $\{\boldsymbol{x}_n\} \in K$. Thus there exists a subsequence $\{\boldsymbol{x}_{n_i}\}$ that converges to some limit in $C$ (since $C$ is closed), call it $\boldsymbol{x}_\star$. So the set of subsequence limits is nonempty.

Claim 1. $\boldsymbol{x}_\star$ is a fixed point, *i.e.*, $\boldsymbol{x}_\star = T(\boldsymbol{x}_\star)$.
Since $\boldsymbol{x}_{n_i} \to \boldsymbol{x}_\star$, by continuity of $T$, $T(\boldsymbol{x}_{n_i}) \to T(\boldsymbol{x}_\star)$, *i.e.*, $\boldsymbol{x}_{n_i+1} \to \boldsymbol{x}' \triangleq T(\boldsymbol{x}_\star)$.
By continuity of $\Psi$, $\Psi(\boldsymbol{x}_{n_i}) \to \Psi(\boldsymbol{x}_\star)$ and $\Psi(\boldsymbol{x}_{n_i+1}) \to \Psi(\boldsymbol{x}')$.
Since $\{n_i\}$ is increasing, $n_{i+1} \geq n_i + 1 > n_i$. Since $T$ is strictly monotonic, $\{\Psi(\boldsymbol{x}_n)\}$ is monotonically nonincreasing.
Thus $\Psi\left(\boldsymbol{x}_{n_{i+1}}\right) \leq \Psi(\boldsymbol{x}_{n_i+1}) \leq \Psi(\boldsymbol{x}_{n_i})$.
Taking the limit as $i \to \infty$ shows that $\lim_{i\to\infty} \Psi(\boldsymbol{x}_{n_i+1}) = \lim_{i\to\infty} \Psi(\boldsymbol{x}_{n_i})$, *i.e.*, $\Psi(\boldsymbol{x}') = \Psi(\boldsymbol{x}_\star)$.

Now if $\boldsymbol{x}_\star$ were not a fixed point of $T$, then by strict monotonicity, since $\boldsymbol{x}' = T(\boldsymbol{x}_\star)$ we would have $\Psi(\boldsymbol{x}') < \Psi(\boldsymbol{x}_\star)$, contradicting the equality just shown. Thus $\boldsymbol{x}_\star \in T_\star$.

Since $\boldsymbol{x}_\star$ was an arbitrary subsequence limit, we have shown $\boxed{\boldsymbol{x}_n} \subseteq T_\star$.

Claim 2. $\Psi(\boldsymbol{x}_n) \to \Psi(\boldsymbol{x}_\star)$.
This follows immediately from the fact $\Psi(\boldsymbol{x}_{n_i}) \to \Psi(\boldsymbol{x}_\star)$ and the fact that $\{\Psi(\boldsymbol{x}_n)\}$ is monotone nonincreasing.
$\forall m \in \mathbb{N}, \exists i \in \mathbb{N}$ s.t. $n_i > m$, and $\Psi(\boldsymbol{x}_{n_i}) \leq \Psi(\boldsymbol{x}_m)$ .

Claim 3. $\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| \to 0$, *i.e.*, $\forall \varepsilon > 0$, $\exists N$ s.t. $n > N \Longrightarrow \|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| < \varepsilon$.
Suppose not, *i.e.*, $\exists \varepsilon > 0$ s.t. $\forall N\ \exists n > N$ s.t. $\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| \geq \varepsilon$.
Then we could recursively construct a subsequence $\{k_i\}$ such that $\|\boldsymbol{x}_{k_i+1} - \boldsymbol{x}_{k_i}\| \geq \varepsilon > 0$.
Since the sequence $\{\boldsymbol{x}_{k_i}\}$ lies in the compact set $K$, it has a convergent subsequence, call it $\{\boldsymbol{x}_{n_i}\}$ for simplicity, that converges to some $\boldsymbol{x} \in K$.
By continuity of $T$, $T(\boldsymbol{x}_{n_i}) \to T(\boldsymbol{x})$, *i.e.* $\boldsymbol{x}_{n_i+1} \to \boldsymbol{x}' \triangleq T(\boldsymbol{x})$.
By Claim 1, $\boldsymbol{x}$ is a fixed point, so in fact $\boldsymbol{x}' = T(\boldsymbol{x}) = \boldsymbol{x}$, so $\boldsymbol{x}_{n_i+1} \to \boldsymbol{x}$.

Since $\boldsymbol{x}_{n_i} \to \boldsymbol{x}$ and $\boldsymbol{x}_{n_i+1} \to \boldsymbol{x}$, there exists some $i \in \mathbb{N}$ such that $\|\boldsymbol{x}_{n_i} - \boldsymbol{x}\| < \varepsilon/2$ and $\|\boldsymbol{x}_{n_i+1} - \boldsymbol{x}\| < \varepsilon/2$.
So by the triangle inequality, for that $i$: $\|\boldsymbol{x}_{n_i+1} - \boldsymbol{x}_{n_i}\| \leq \|\boldsymbol{x}_{n_i} - \boldsymbol{x}\| + \|\boldsymbol{x}_{n_i+1} - \boldsymbol{x}\| < \varepsilon$, a contradiction.
Thus $\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| \to 0$.

Claim 4. $\boxed{\boldsymbol{x}_n}$ is connected.
This follows immediately from **Ostrowki's convergence theorem**. □

One tries to choose $T$ such that the fixed points of $T$ are the minimizers of $\Psi$.

**Corollary**. If $C = \{\boldsymbol{x} \in \mathcal{X} : \Psi(\boldsymbol{x}) \leq \Psi(\boldsymbol{x}_0)\}$ is compact and $T$ is continuous and strictly monotonic on $C$ with respect to a continuous function $\Psi$ with a unique minimizer $\boldsymbol{x}_\star$, and if $T_\star = \{\boldsymbol{x}_\star\}$, then the algorithm (M-3) yields a sequence $\{\boldsymbol{x}_n\}$ that converges to the unique minimizer.

***p. 111 of Meyer claims that if $\{x_n\}$ does not converge, then "there exists at least two accumulation points"***

Extra credit (= 40 homework points) problem. Show that Meyer's definition of u.s.c. is equivalent to our definition (or give a counter-example).

Example. ***overhead***

Separable paraboloidal surrogate algorithm for ordinary least squares estimation [10–13].

Consider the problem of finding the minimizer of $\Psi(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|_2^2$, where $\boldsymbol{y} \in \mathbb{R}^m$, $\boldsymbol{x} \in \mathbb{R}^p$, and $A \in \mathbb{R}^{m \times p}$. Of course there is a simple analytical solution for the *unconstrained* minimizer of $\Psi(\boldsymbol{x})$, but what if we want $\min_{\boldsymbol{x} \geq \mathbf{0}} \Psi(\boldsymbol{x})$?

To derive an algorithm for the constrained case, note that

$$\Psi(\boldsymbol{x}) = \sum_{i=1}^m h_i([A\boldsymbol{x}]_i), \text{ where } h_i(l) = \frac{1}{2}(y_i - l)^2 \text{ and } [A\boldsymbol{x}]_i = \sum_{j=1}^p a_{ij} x_j.$$

Now (this clever trick is due to De Pierro [14]) consider

$$[A\boldsymbol{x}]_i = \sum_{j=1}^p a_{ij} x_j = \sum_{j=1}^p \pi_{ij} \left( \frac{a_{ij}}{\pi_{ij}} (x_j - x_j^n) + [A\boldsymbol{x}^n]_i \right),$$

where $\pi_{ij} = \frac{|a_{ij}|}{a_i} \geq 0$ and $a_i \triangleq \sum_j |a_{ij}| > 0$. (If $a_i = 0$ then that row of $A$ is pointless and should be eliminated at the outset.) Note that $\sum_{j=1}^p \pi_{ij} = 1$.

Since each $h_i$ is a **convex function**:

$$h_i([A\boldsymbol{x}]_i) = h_i\left( \sum_{j=1}^p \pi_{ij} \left( \frac{a_{ij}}{\pi_{ij}} (x_j - x_j^n) + [A\boldsymbol{x}^n]_i \right) \right) \leq \sum_{j=1}^p \pi_{ij} h_i\left( \frac{a_{ij}}{\pi_{ij}} (x_j - x_j^n) + [A\boldsymbol{x}^n]_i \right).$$

Thus

$$\Psi(\boldsymbol{x}) \leq g(\boldsymbol{x}; \boldsymbol{x}_n) \triangleq \sum_{i=1}^m \sum_{j=1}^p \pi_{ij} h_i\left( \frac{a_{ij}}{\pi_{ij}} (x_j - x_j^n) + [A\boldsymbol{x}^n]_i \right),$$

but $\Psi(\boldsymbol{x}_n) = g(\boldsymbol{x}_n; \boldsymbol{x}_n)$. ***(Picture of surrogate)***

Consider the following iterative algorithm:

$$\boldsymbol{x}_{n+1} = T(\boldsymbol{x}_n) = \arg\min_{\boldsymbol{x} \geq \mathbf{0}} g(\boldsymbol{x}; \boldsymbol{x}_n).$$

This algorithm is monotonic since $\Psi(\boldsymbol{x}_{n+1}) \leq g(\boldsymbol{x}_{n+1}; \boldsymbol{x}_n) \leq g(\boldsymbol{x}_n; \boldsymbol{x}_n) = \Psi(\boldsymbol{x}_n)$. Is it strictly monotone? Continuous?

$$\frac{\partial}{\partial x_j} g(\boldsymbol{x}; \boldsymbol{x}_n) = \sum_i a_{ij} \dot{h}_i\left( \frac{a_{ij}}{\pi_{ij}} (x_j - x_j^n) + [A\boldsymbol{x}^n]_i \right) = \sum_i a_{ij} \left[ y_i - \frac{a_{ij}}{\pi_{ij}} (x_j - x_j^n) - [A\boldsymbol{x}^n]_i \right].$$

Thus, equating to zero yields the iteration

$$\boxed{x_j^{n+1} = \left[ x_j^n + \frac{\sum_i a_{ij}(y_i - [A\boldsymbol{x}^n]_i)}{\sum_i |a_{ij}|\, a_i} \right]_+}, \qquad \boxed{\boldsymbol{x}^{n+1} = \left[ \boldsymbol{x}_n + \mathsf{diag}\left\{ \frac{1}{\sum_i |a_{ij}|\, a_i} \right\} A'(\boldsymbol{y} - A\boldsymbol{x}_n) \right]_+},$$

where $[x]_+$ denotes $x$ if it is nonnegative and zero otherwise.

Continuity is obvious. With some algebra one can show that $T$ is strictly monotone too.

One can show that $\boldsymbol{x}_\star$ is a fixed point of $T$ iff $\boldsymbol{x}_\star$ satisfies the **Karush Kuhn Tucker** conditions for the minimization of $\Psi$.

One can show that if $A$ has full column rank, then there is a unique minimizer that is the only fixed point of $T$ and the above algorithm converges to that minimizer [15, 16].

# Bibliography


[1] P. Enflo. A counterexample to the approximation problem in Banach spaces. *Acta Math*, 130:309–17, 1973.

[2] I. J. Maddox. *Elements of functional analysis*. Cambridge, 2 edition, 1988.

[3] A. W. Naylor and G. R. Sell. *Linear operator theory in engineering and science*. Springer-Verlag, New York, 2 edition, 1982.

[4] D. G. Luenberger. *Optimization by vector space methods*. Wiley, New York, 1969.

[5] J. Schauder. Zur theorie stetiger abbildungen in funktionenrumen. *Math. Zeitsch.*, 26:47–65, 1927.

[6] P. P. Vaidyanathan. Generalizations of the sampling theorem: Seven decades after Nyquist. *IEEE Tr. Circ. Sys. I, Fundamental theory and applications*, 48(9):1094–109, September 2001.

[7] A. M. Ostrowski. *Solution of equations in Euclidian and Banach spaces*. Academic, 3 edition, 1973.

[8] R. R. Meyer. Sufficient conditions for the convergence of monotonic mathematical programming algorithms. *J. Comput. System. Sci.*, 12(1):108–21, 1976.

[9] M. Rosenlicht. *Introduction to analysis*. Dover, New York, 1985.

[10] A. R. De Pierro. On the relation between the ISRA and the EM algorithm for positron emission tomography. *IEEE Tr. Med. Im.*, 12(2):328–33, June 1993.

[11] A. R. De Pierro. On the convergence of the iterative image space reconstruction algorithm for volume ECT. *IEEE Tr. Med. Im.*, 6(2):174–175, June 1987.

[12] A. R. De Pierro. Unified approach to regularized maximum likelihood estimation in computed tomography. In *Proc. SPIE 3171, Comp. Exper. and Num. Meth. for Solving Ill-Posed Inv. Imaging Problems: Med. and Nonmed. Appl.*, pages 218–23, 1997.

[13] J. A. Fessler. Grouped coordinate descent algorithms for robust edge-preserving image restoration. In *Proc. SPIE 3170, Im. Recon. and Restor. II*, pages 184–94, 1997.

[14] A. R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Tr. Med. Im.*, 14(1):132–137, March 1995.

[15] J. A. Fessler and A. O. Hero. Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. *IEEE Tr. Im. Proc.*, 4(10):1417–29, October 1995.

[16] M. W. Jacobson and J. A. Fessler. Properties of optimization transfer algorithms on convex feasible sets. *SIAM J. Optim.*, 2003. Submitted.

[17] F. Deutsch. The convexity of Chebyshev sets in Hilbert space. In A. Yanushauskas Th. M. Rassias, H. M. Srivastava, editor, *Topics in polynomials of one and several variables and their applications*, pages 143–50. World Sci. Publishing, River Edge, NJ, 1993.

[18] M. Jiang. On Johnson's example of a nonconvex Chebyshev set. *J. Approx. Theory*, 74(2):152–8, August 1993.

[19] V. S. Balaganskii and L. P. Vlasov. The problem of convexity of Chebyshev sets. *Russian Mathematical Surveys*, 51(6):1127–90, November 1996.

[20] V. Kanellopoulos. On the convexity of the weakly compact Chebyshev sets in Banach spaces. *Israel Journal of Mathematics*, 117:61–9, 2000.

[21] A. R. Alimov. On the structure of the complements of Chebyshev sets. *Functional Analysis and Its Applications*, 35(3):176–82, July 2001.

[22] P. L. Combettes and H. J. Trussell. Method of successive projections for finding a common point of sets in metric spaces. *J. Optim. Theory Appl.*, 67(3):487–507, December 1990.

[23] M. Vetterli and J. Kovacevic. *Wavelets and subband coding*. Prentice-Hall, New York, 1995.

[24] D. C. Youla and H. Webb. Image restoration by the method of convex projections: Part I—Theory. *IEEE Tr. Med. Im.*, 1(2):81–94, October 1982.