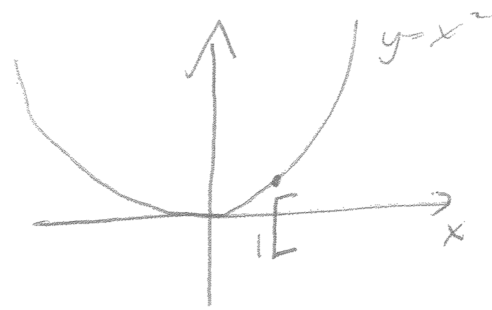
ALGORITHMS FOR NONLINEAR PROGRAMMING

Eg:

$$\min \; x^2$$
$$\text{s.t.} \; x \geq 1$$



Optimal solution $\bar{x} = 1$

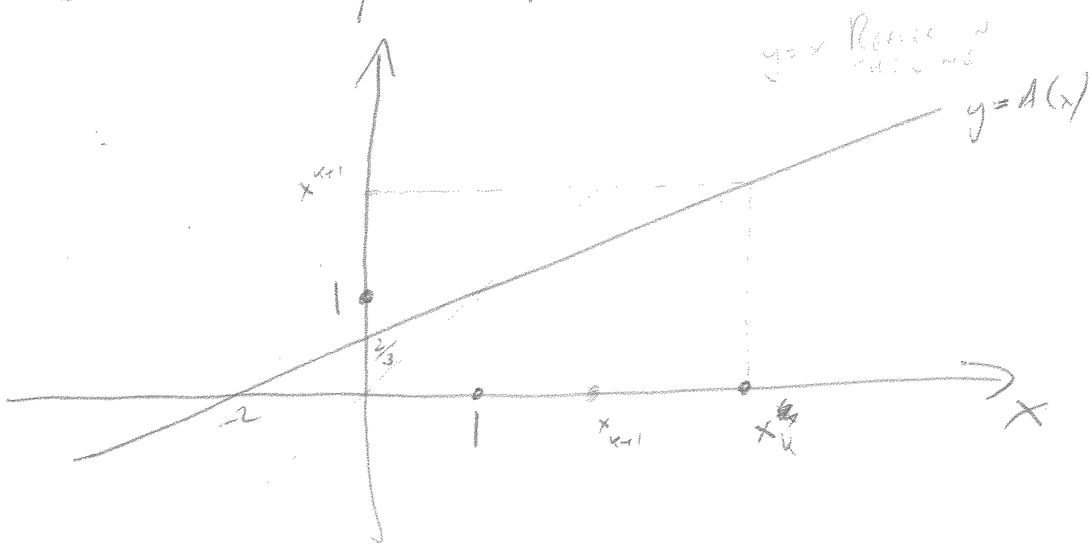
Consider the algorithm: $A(x) = \frac{1}{3}(x+2)$ for any point x;
$A(1) = 1$.
For example, if we start with $x_1 = 5$, get:

$$x_1 = 5, \quad x_2 = \tfrac{1}{3}(x_1+2) = \tfrac{7}{3}, \quad x_3 = \tfrac{1}{3}(x_2+2) = \tfrac{1}{3}\left(\tfrac{13}{3}\right) = \tfrac{13}{9}$$
$$x_4 = \tfrac{1}{3}\left(\tfrac{13}{9}+2\right) = \tfrac{31}{27} \qquad x_5 = \tfrac{1}{3}\left(\tfrac{31}{27}+2\right) = \tfrac{85}{81} \ldots$$

converges to $\bar{x} = 1$.

~~Can also have POINT-TO-SET algorithms: For each point x, define $A(x)$ to be a SET of points.~~

Draw a picture of this ALGORITHMIC MAP:

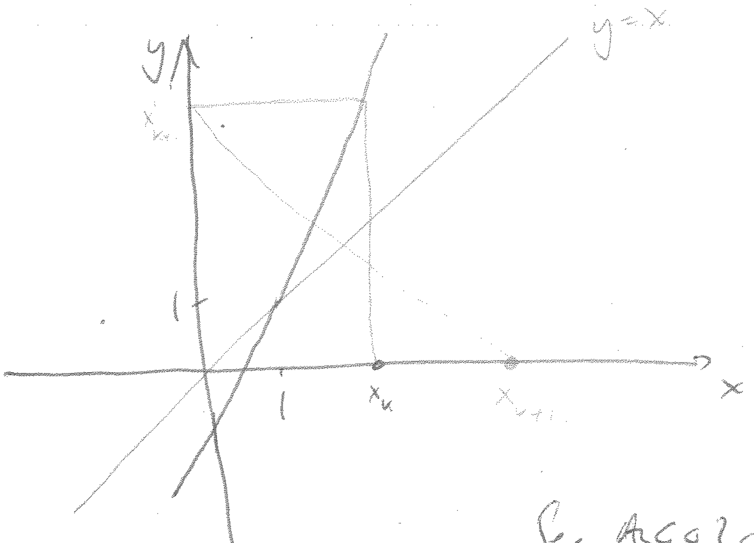


Transformation of x_k into x_{k+1} is an ITERATION.

Note: $A: X \to X$, where X is domain of the problem.

Consider the algorithm: $A(x) = 2(x - \frac{1}{2})$ for any point x.
$A(1) = 1$.
For example, start with $x_1 = 2$, get:

$x_1 = 2$, $x_2 = 2(2-\frac{1}{2}) = 3$, $x_3 = 2(3-\frac{1}{2}) = 5$, $x_4 = 2(5-\frac{1}{2}) = 9$, ...
Diverges.



See ALGO 2a for motivation for Newton's method.

Eg: Newton's method for unconstrained problems:

$$\min f(x) \qquad f: \mathbb{R} \to \mathbb{R}.$$

$$A(x) = x - \frac{f'(x)}{f''(x)}$$

Also: Newton for $f(x) = -x$ converges to max x!

Newton for $f(x) = x^4$
$x^{k+1} = x^k - \frac{1}{3} x^k$
$= \frac{2}{3} x^k$.

Also: $f(x) = x^{4/3}$: Diverges.

Eg: min $x^2 + 3x$ ~~x_5~~ Then $f'(x) = 2x+3$, $f''(x) = 2$.

~~With~~

With $x_1 = 1$: $\quad x_2 = x_1 - \frac{2x_1+3}{2} = 1 - \frac{5}{2} = -\frac{3}{2}$, optimal.

Newton's method will solve a quadratic problem in one step.
i.e. find the stationary point

ALGO 2a

min $f(x)$ $\quad f: \mathbb{R} \to \mathbb{R}$

$$f(x) \approx f(\bar{x}) + f'(\bar{x})(x-\bar{x}) + \tfrac{1}{2} f''(\bar{x})(x-\bar{x})^2 =: q(x)$$

Minimize the quadratic approximation:

$$\frac{dq}{dx} = \cancel{f'(x)} + f''(\bar{x})(x-\bar{x})$$

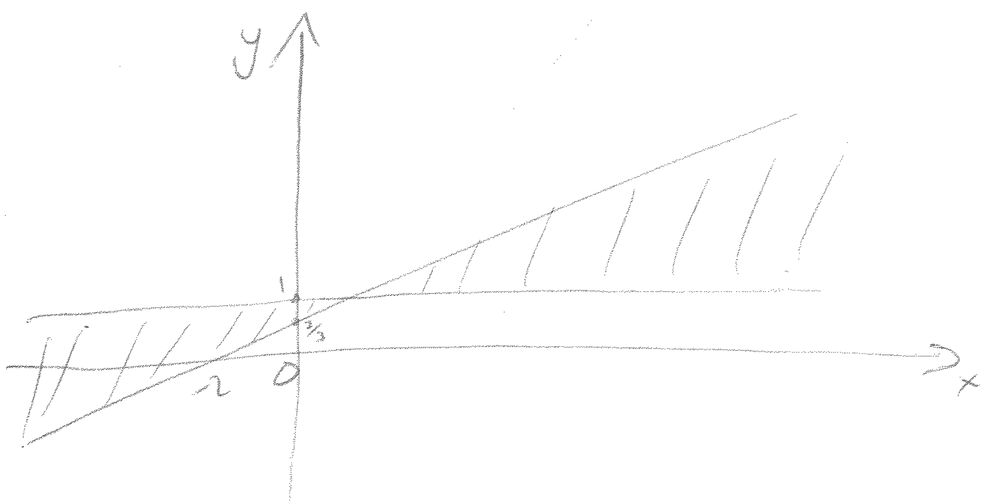
So take $x - \bar{x} = -\dfrac{f'(\bar{x})}{f''(\bar{x})}$,

or $x = \bar{x} - \dfrac{f'(\bar{x})}{f''(\bar{x})}$.

Can also have POINT-TO-SET algorithms:
For each point x, define $A(x)$ to be a SET of points.

e.g. $\min \ x^2$
s.t. $x \geq 1$.

$$A(x) = \begin{cases} \left[1, \tfrac{1}{3}(x+2)\right] & \text{if } x \geq 1 \\ \left[\tfrac{1}{3}(x+2), 1\right] & \text{if } x \cancel{\geq} \leq 1. \end{cases}$$



One possible sequence is:

$x_1 = 5, \quad x_2 = 2 \ \left(\in \left[1, \tfrac{7}{3}\right]\right)$
$x_3 = 1.1 \left(\in \left[1, \tfrac{4}{3}\right]\right), \ldots$

THE SOLUTION SET AND CONVERGENCE OF ALGORITHMS

Solve $\min f(x)$

s.t. $x \in S$.

Would like an algorithm generating a sequence of points converging to an optimal solution.

This is optimistic. So usually look to converge to some set Ω.

Possible choices for solution set Ω include:

1. $\Omega = \{\bar{x} : \bar{x}$ is a local optimal solution$\}$.

2. $\Omega = \{\bar{x} : \bar{x} \in S, f(\bar{x}) \leq b\}$ for some acceptable value b.

3. $\Omega = \{\bar{x} : \bar{x} \in S, f(\bar{x}) < LB + \varepsilon\}$ for some tolerance ε, and LB is a lower bound on optimal value.
   E.g.: LB comes from Lagrangian dual.

4. $\Omega = \{\bar{x} : \bar{x} \in S, f(\bar{x}) \leq v^* + \varepsilon\}$ for some tolerance ε, where v^* is known global minimum value.

5. $\Omega = \{\bar{x} : \bar{x} \in S, \bar{x}$ satisfies KKT optimality conditions$\}$.

6. $\Omega = \{\bar{x} : \bar{x} \in S, \bar{x}$ satisfies Fritz-John optimality conditions$\}$.

An algorithmic map $A : X \to X$ CONVERGES over $Y \subseteq X$ if, when started from any $x_1 \in Y$, the limit of any convergent subsequence $x_1, x_2, \ldots$ belongs to Ω.

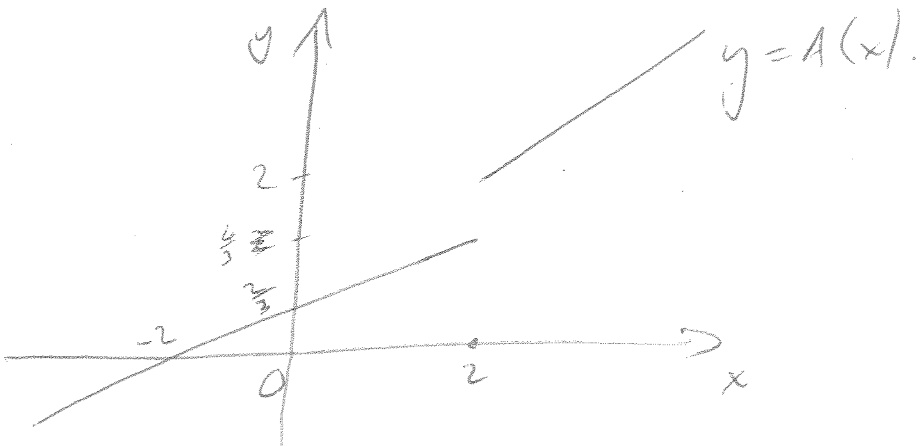
Closed Maps

Eg: min x^2 Optimal soln $\bar{x} = 1$.
s.t. $x \geq 1$.   Let $\Omega = \{1\}$.

Consider the algorithm:

$$A(x) = \begin{cases} \frac{1}{3}(x+2) & \text{if } x \leq 2 \\ \frac{1}{2}(x+2) & \text{if } x > 2. \end{cases}$$

(≤ 2, > 2 differ from book.)



If $x_1 > 2$: converge to $x = 2$

If $1 \leq x_1 \leq 2$: converge to $\bar{x} = 1$

This example and the earlier convergent examples satisfy:

(1) Given feasible x_k, then x_{k+1} is feasible

(2) For any x not in solution set, $f(x_{k+1}) < f(x_k)$

(3) Given x in solution set, (that is, $x_k = 1$), then $x_{k+1} = 1$.

So these are not enough to guarantee convergence!

DEFN: Let X and Y be nonempty closed sets. Let $A: X \to Y$ be a point-to-set (or point-to-point) map. The map is CLOSED at $x \in X$ if:

given a sequence $x_k \in X$, $x_k \to \bar{x}$
given a sequence $y_k \in A(x_k)$, $y_k \to \bar{y}$

then $\bar{y} \in A(\bar{x})$.

The map is closed on $Z \subseteq X$ if it is closed at each point in Z.

Eg: the last example is not closed, at $x = 2$.

Because: $x_k = 2 + 1/k$
Then $y_k = \frac{1}{2}(x+2) = 2 + \frac{1}{2k}$

$x_k \to \bar{x} = 2$, but $A(2) = 4/3 \neq \lim y_k$.

Eg: $\Omega = [0,1]$, $\quad Z = [-0.99, 1.99]$.

$$A(x) = \begin{cases} 1+x^2 & \text{if } x < 0 \\ -(x-1)^2 & \text{if } x > 1 \\ 1-x & \text{if } x \in [0,1]. \end{cases}$$

Then get, eg:

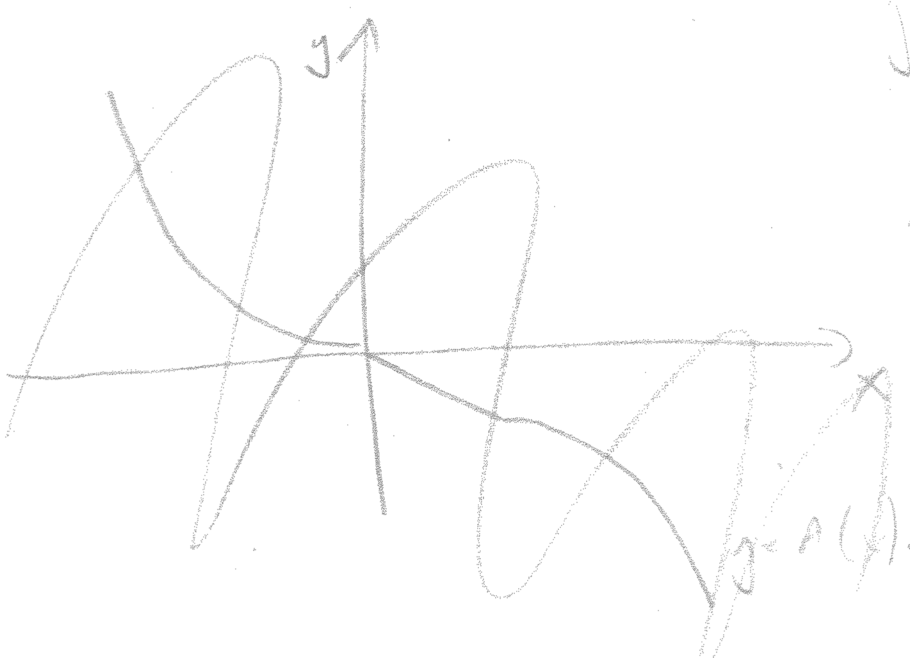
$x_1 = 1.5$, $x_2 = -0.25$, $x_3 = 1\frac{1}{16}$, $x_4 = -\frac{1}{256}$, $x_5 = 1+\frac{1}{65536}$, ...

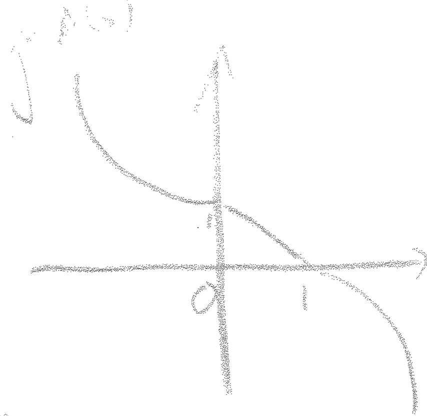
This is closed:

If $x_k \to \bar{x}$ and $y_k \in A(x_k) \to \bar{y}$ then $A(\bar{x}) = \bar{y}$.

Notice the sequence has <u>two</u> accumulation points, 0 and 1.

Graph of algorithmic map:





Note that this map is not ~~closed~~ monotonic if $\Omega = \{0.5\}$,

~~since~~ and $\alpha = z(x) = (x-0.5)^2$.

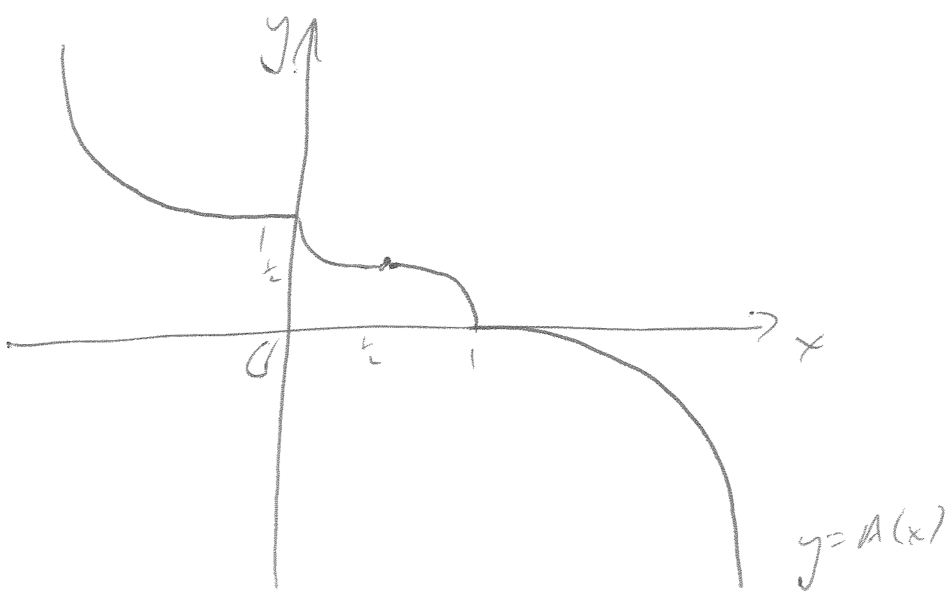
Since: $A(0) = 1$, and $\alpha(0) = \alpha(1)$.

$A(\frac{1}{4}) = \frac{3}{4}$, and $\alpha(\frac{1}{4}) = \alpha(\frac{3}{4})$.

Changing it to:

$$A(x) = \begin{cases} 1 + x^2 & \text{if } x < 0 \\ -(x-1)^2 & \text{if } x > 1 \\ \frac{1}{2} + (x-\frac{1}{2})^2 & \text{if } \frac{1}{2} \le x \le 1 \\ \frac{1}{2} - (x-\frac{1}{2})^2 & \text{if } 0 \le x \le \frac{1}{2} \end{cases}$$

still keeps it ~~nonmonot~~ nonmonotonic, but only at $x = 0$ and $x = 1$.



Get convergence if $x_1 \in (0,1)$.

Before proving theorem on next page:

descent function

Possible choices for α:

- Objective function $f(x)$
- For unconstrained problems: $\|\nabla f(x)\|$
- A MERIT FUNCTION:

For $\min f(x)$
s.t. $g_i(x) \leq 0$,

take, eg, $\alpha(x) = f(x) + \sum_i \max\{0, g_i(x)\}$,

so we penalise violated constraints.

Or: $\alpha(x) = f(x) + \sum_i (\max\{0, g_i(x)\})^2$

– then differentiable ~~at~~ when $g_i(x) = 0$, provided g_i differentiable.

Convergence Theorem

$X \subseteq \mathbb{R}^n$, $X \neq \emptyset$. Let $\Omega \subseteq X$ be solution set, $\Omega \neq \emptyset$.

Let $A: X \to X$ ~~by~~ be point-to-set map. Given $x_1 \in X$, the sequence $\{x_k\}$ is generated iteratively as:

If $x_k \in \Omega$, then STOP;
Else, let $x_{k+1} \in A(x_k)$, replace k by $k+1$, repeat.

Suppose $\{x_k\}$ is contained in a compact subset of X.

Assume there exists a continuous function α (the DESCENT FUNCTION), such that $\alpha(y) < \alpha(x)$ if $x \notin \Omega$ and $y \in A(x)$.

If A is closed over the complement of Ω then either the algorithm stops in a finite number of steps with a point in Ω, or it generates ~~a~~ the infinite sequence $\{x_k\}$ such that

1. Every convergent subsequence of $\{x_k\}$ has a limit in Ω.

2. $\alpha(x_k) \to \alpha(x)$ for some $x \in \Omega$.

Proof Suppose an infinite sequence $\{x_k\}$ is generated.

Let $\{x_k\}_K$ be any convergent subsequence with limit $\bar{x} \in X$.

Want to show: $\lim \{x_k\}_K = \bar{x} \in \Omega$, $\alpha(x_k) \to \alpha(\bar{x})$ for some $\bar{x} \in \Omega$.

Since α is continuous, we have $\alpha(x_k) \to \alpha(\bar{x})$.

So, given $\varepsilon > 0$, for k large enough, say $k > k_0$: $\alpha(x_k) < \alpha(\bar{x}) + \varepsilon$

α is descent function, so $\alpha(x_k) \geq \alpha(\bar{x})$.

Thus, $\alpha(\bar{x}) \leq \alpha(x_k) < \alpha(\bar{x}) + \varepsilon \quad \forall k > k_0 \quad (*)$.

True for all $k > k_0$, not just $k \in$ subsequence K.

Need to show $\bar{x} \in \Omega$.

For ~~$\bar{x}_k$~~ our subsequence $\{x_k\}_K$, consider the ~~im~~ images of these points under the algorithmic map:

$\{x_{k+1}\}_K$.

This ~~sub~~sequence is contained in a compact set, so it has a convergent subsequence $\{x_{k+1}\}_{\bar{K}}$, with limit $\hat{x} \in$ ~~X~~ X.

Must have ~~lim~~ $\alpha(\hat{x}) = \alpha(\bar{x})$, since have continuous α, and from relationship $(*)$.

ALGO 11

~~Since A is closed, and since $x_k \to \bar{x}$ and~~

We have: A is closed

$\{x_k\}_K \to \bar{x}$

$x_{k+1} \in A(x_k)$ and $x_{k+1} \to \hat{x}$

So we have $\hat{x} \in A(\bar{x})$.

Thus, ~~$\alpha(\hat{x}) < \alpha(\bar{x})$, but this contradicts $(*)$.~~
if $\bar{x} \notin \Omega$ then $\alpha(\hat{x}) < \alpha(\bar{x})$, but this contradicts $(*)$.

So we must have $\lim \{x_k\}_K = \bar{x} \in \Omega$.

∎

COROLLARY Under the assumptions of the theorem, if Ω is the singleton $\bar{x}$ then the whole sequence $\{x_k\}$ converges to $\bar{x}$.

PROOF. Suppose $\exists$ subsequence K with $\|x_k - \bar{x}\| > \varepsilon$ for $k \in K$. $(+)$

$\exists$ subsequence $K' \subseteq K$ such that $\{x_k\}_{K'}$ has limit x', and $x' \in \Omega$.

But $\Omega = \{\bar{x}\}$, so $x' = \bar{x}$. Thus, $\lim\limits_{k \in K'} x_k = \bar{x}$, contradicting $(+)$

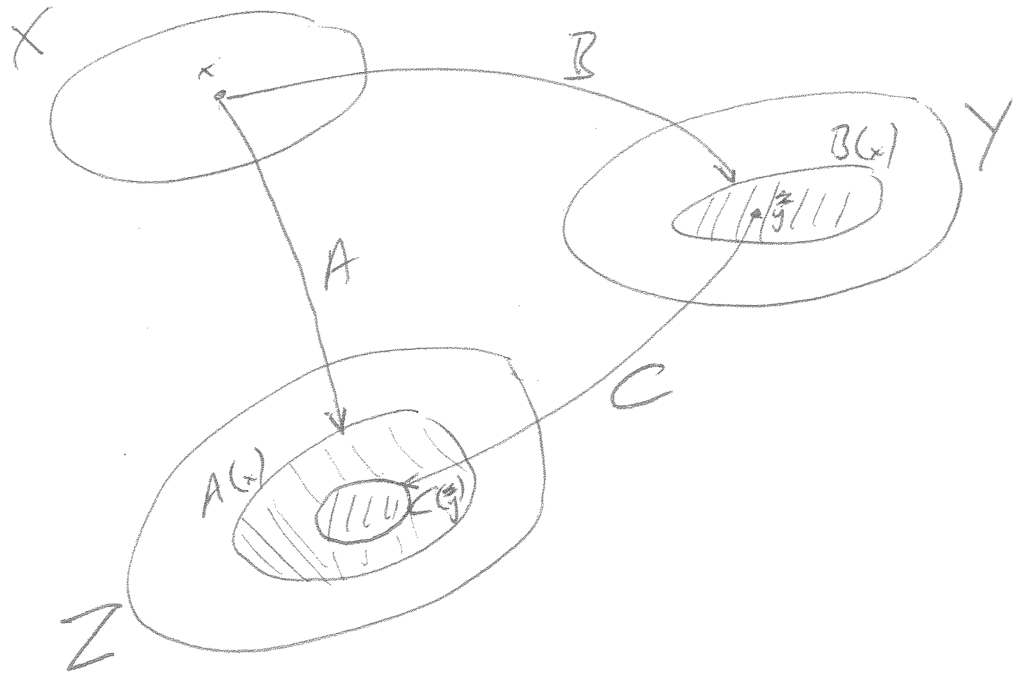
∎

COMPOSITION OF MAPPINGS

Eg: Choose direction ← one map
then choose step length ← another map.

Eg: simplex algorithm:
Choose incoming variable ⟷ choose a particular edge to move along

then: choose leaving variable ⟷ choose step length along the edge.

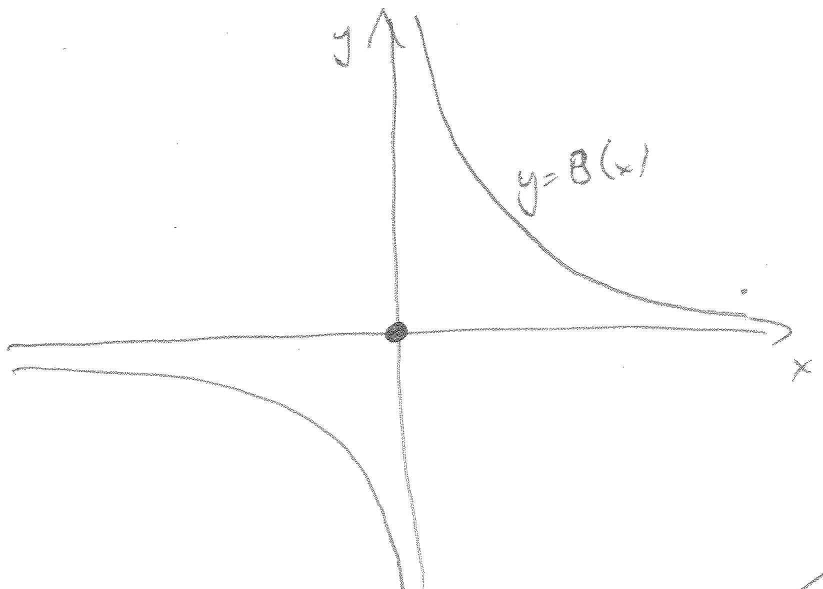


Each point in X is mapped to a set in Z.

First map each point in X to a set in Y.
Then each point y in this set is mapped to its own set $C(y)$ in Z.
We get $A(x) = \bigcup_{y \in B(x)} C(y)$.

Eg: $x \in \mathbb{R}$.

$$B(x) = \begin{cases} 1/x & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

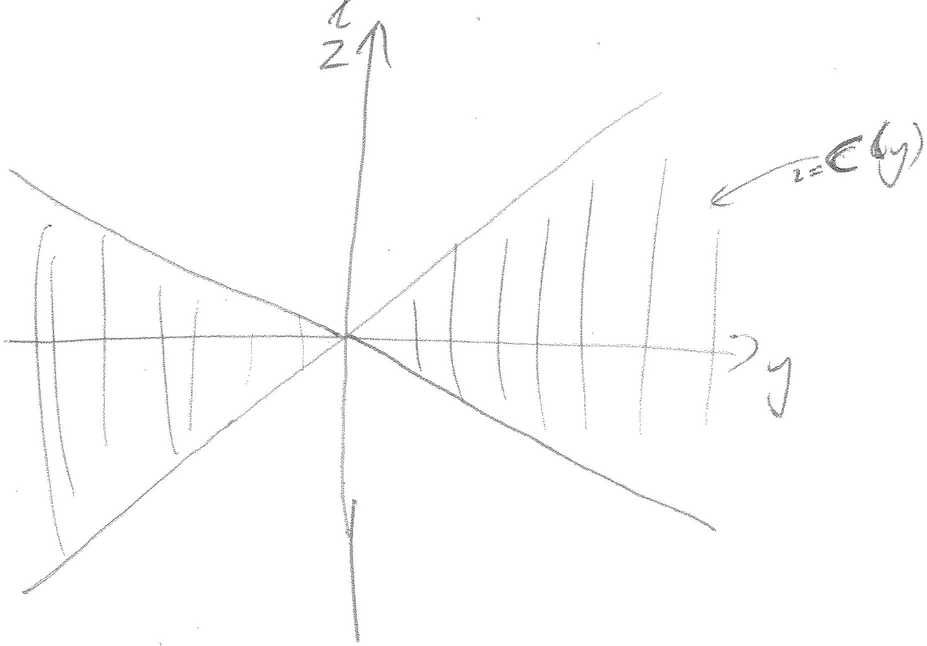


This is closed:
If $x \neq 0$ then any sequence $x_k \to x$ will also have $B(x_k) \to B($

If $x = 0$ then any sequence $x_k \to x$, ~~will have~~ $x_k \neq x$, will have $B(x_k)$ diverging.
The only possible convergent subsequences K in $B(x_k)$ with $x_k \to x$ are those with $x_k = 0$ $\forall k \in K$

So vacuously closed at $x=0$.

$$C(y) = \{ z : |z| \leq |y| \} \quad \text{for any } y \in \mathbb{R}.$$



This is closed:
If $y_k \to y$ and $z_k \to z$ and $z_k \in C(y_k)$ then we must have $z \in C(y)$.

Now look at the composition of the maps:

$$A = CB \qquad \text{(ie, } B \text{ followed by } C\text{).}$$

We claim A is not closed at $x = 0$:

Consider $x_k = \frac{1}{k} \to 0$ as $k \to \infty$.
Then $A(x_k) = C(B(x_k)) = C(k) = [-k, k]$

Thus, with $y_k = 1\ \forall k$, ~~is an element of $A(x_k)\ \forall k$.~~
we have $y_k \in A(x_k)\ \forall k$
and $y_k \to 1 \notin A(0) = C(B(0)) = C(0)$
$= \{0\}$.

What messed this up was that there were no convergent series $x_k \to 0$ with $y_k \in B(x_k)$ and y_k convergent.

We have:

Theorem Let $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^l$, $Z \subseteq \mathbb{R}^m$.
Let $B : X \to Y$, $C : Y \to Z$, $A = CB : X \to Z$.
Assume B is closed at $\bar{x}$ and C is closed ~~at~~ on $B(\bar{x})$.
(*) { Assume that if $x_k \to \bar{x}$ ~~then~~ ~~there exists~~ and $y_k \in B(x_k)$ then there is a convergent subsequence of $\{y_k\}$.
Then A is closed at $\bar{x}$.

(No proof. See text for proof.)

Corollary Replace (*) by assumption that ~~If in addition~~ Y is compact then A is closed.

Corollary Replace (*) by assumption that B is continuous. Then A is closed.

Convergence of Algos with Composite Maps.

Eg: $A = BC$.

B closed : eg, find direction

C may not be closed. Only requirement is that it does not increase the merit function. Eg: line search, with some tolerance for accepting step.

Theorem $\Omega \subseteq X \subseteq \mathbb{R}^n$, $\Omega \neq \phi$, Ω is solution set. $\alpha : \mathbb{R}^n \to \mathbb{R}$, continuous

$C : X \to X$ satisfies: given $x \in X$ then $\alpha(y) \leq \alpha(x)$ for $y \in C(x)$.

$B : X \to X$ point-to-set, closed over complement of Ω, $\alpha(y) < \alpha(x)$ for each $y \in B(x)$, if $x \notin \Omega$.

Let $A = CB$. Suppose $x_1 \in X$ is given.

Define $\{x_k\}$ by: $x_{k+1} \begin{cases} \in A(x_k) & \text{if } x_k \notin \Omega \\ \text{else STOP.} \end{cases}$

Assume $\Lambda = \{x : \alpha(x) \leq \alpha(x_1)\}$ is compact.

Then either the algorithm ~~converges~~ stops in a finite number of steps with a point in Ω or all accumulation points of $\{x_k\}$ belong to Ω.

(Note: the example on page ALGO 15 does not have a function α. If we try $\alpha = |x|$, then B does not satisfy $\alpha(y) < \alpha(x)$ for each $y \in B(x)$.)

No proof. Proof in text, is similar to theorem we proved last time, on p. ALGO 9 ☺.

Terminating the Algorithm:

Possible stopping rules:

1. $\|x_{k+N} - x_k\| < \varepsilon$:

   Alg is stopped if N steps result in only a small change in the iterate.

2. $\dfrac{\|x_{k+1} - x_k\|}{\max\{1, \|x_k\|\}} < \varepsilon$:

   Algo is stopped if there is only a small relative change in the iterate. The denominator serves to scale the measure: If $\|x_k\|$ is large, then we should tolerate larger changes in x.
   If $\|x_k\|$ is small, then just look at $\|x_{k+1} - x_k\|$.

3. $\alpha(x_k) - \alpha(x_{k+N}) < \varepsilon$:

   Algorithm does not make much progress in the descent function in N steps.

4. $\dfrac{\alpha(x_k) - \alpha(x_{k+1})}{\max\{1, |\alpha(x_k)|\}} < \varepsilon$:

   Relative improvement is small.

5. $\alpha(x_k) < C$ for some constant C.

   Used in Karmarkar's original LP algorithm, where he took

$$\alpha(x) = n \log c^T x - \sum_{i=1}^{n} \log(x_i),$$

and where the LP was manipulated to have an optimal value of 0.
Here, $\alpha(x) \to -\infty$ as $x \to$ optimal soln.

6. $\alpha(x_k) - \alpha(\bar{x}) < \varepsilon$, where $\bar{x}$ belongs to Ω.

Suitable if $\alpha(\bar{x})$ known beforehand, eg, in unconstrained optimization may take $\alpha(x) = \|Df(x)\|$, so $\alpha(\bar{x}) = 0$.

Aug 18

Minimizing along independent directions

$$\min \; f(x) \quad \text{s.t.} \; x \in \mathbb{R}^n.$$

Let $\{d_1, d_2, \ldots, d_n\}$ be a basis for $\mathbb{R}^n$ with $\|d_i\| = 1 \; \forall i$.

0. Start with an iterate x^0, with $k=0$

~~b.~~

1. If x^k meets termination criterion, STOP.

2. $\bar{x} = x^k$. for $i = 1, \ldots, n$

   $\bar{\alpha}_i = \arg\min \{ f(\bar{x} + \alpha_i d_i) \}$ (Note: α_i may be negative.)

   $\bar{x} = \bar{x} + \bar{\alpha}_i d_i$

   loop

3. $x^{k+1} = \bar{x}$. $k = k+1$. Return to Step (1).

Perhaps simplest choice for d_i are the coordinate axes.

Get "coordinate descent" algorithms. — Useful when there is little information about f available, e.g. hard to get gradient, Hessian matrix.

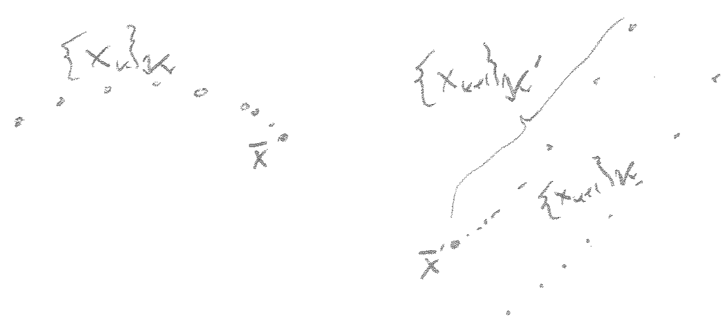
Theorem Consider the algorithm given above. Assume f is differentiable and that the minimum of f along any line in $\mathbb{R}^n$ is unique. If the sequence $\{x_k\}$ is contained in a compact subset of $\mathbb{R}^n$ then each accumulation point x of the sequence $\{x_k\}$ must satisfy

$\nabla f(x_k) = 0$.

Proof. Let $\bar{x}$ be an accumulation point of the infinite subsequence $\{x_k\}_K$.

Need to show $\nabla f(\bar{x}) = 0$.

Consider $\{x_{k+1}\}_K$, ie the next iterate after the subsequence. This is contained in a compact subset of $\mathbb{R}^n$, so it has a convergent subsequence $\{x_{k+1}\}_{K'}$, with limit $\bar{x}'$



We show that taking one step of the algorithm from $\bar{x}$ results in $\bar{x}'$.

Let $D = [d_1, \ldots, d_n]$.

Then $x^*_{k+1} = x_k + \sum_{j=1}^{n} \lambda_{jk} d_j = x_k + D\lambda_k$

Move this here.

Let $y_{0k} = x_k$, $y_{jk} = y_{j-1,k} + \lambda_{jk} d_j$,

so $x_{k+1} = y_{nk}$.

By definition of algorithm, y_{jk} solves $\min\{y_{j-1,k} + \lambda d_j\}$,

~~so $f(y_{jk} + \lambda_{jk} d_j) = f$~~

so $f(y_{jk}) = f(y_{j-1,k} + \lambda_{jk} d_j) \leq f(y_{j-1,k} + \lambda d_j)$ for any λ.

For $k \in K'$, we have $x_k \to \bar{x}$, $x_{k+1} \to \bar{x}'$.

Also, $\cancel{\frac{\lambda_k}{k+1}} = D^{-1}(x_{k+1} - x_k) \longrightarrow \lambda$, since D is invertible.

Thus, $\bar{x}' = \bar{x} + D\lambda$.

We also have $y_{jk} \to y_j$, say,

so $f(y_j) \leq f(y_{j-1} + \lambda d_j)$ for any λ, by continuity of f.

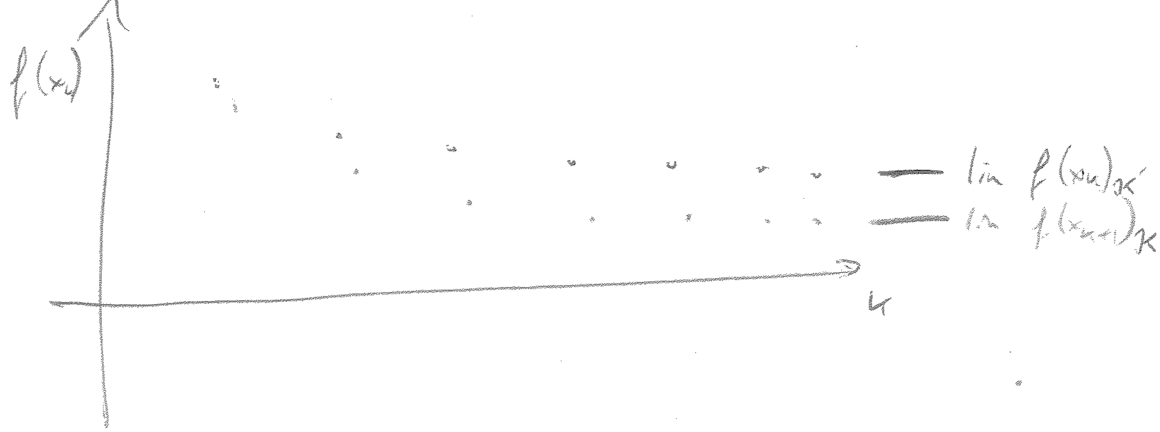
So ~~taking~~ ... step from $\bar{x}$ gives $\bar{x}'$ with intermediate steps y_j.

Alg 021.

By defn of y_{jk}, must have $f(\bar{x}') \leq f(\bar{x})$.

Can we have $f(\bar{x}') < f(\bar{x})$?

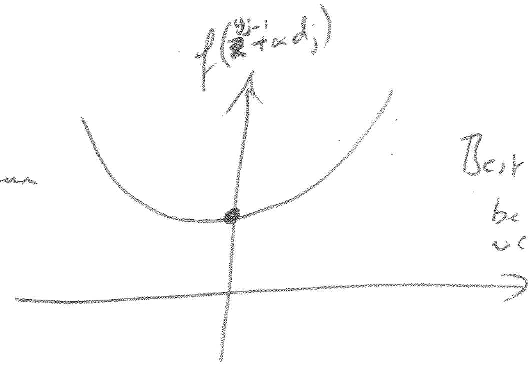
No, because $f(x_k)$ is monotonically decreasing.



So $f(\bar{x}') = f(\bar{x})$.

~~But now $Df($~~

Since there exists a unique minimum along each direction d_j, must have $\bar{x}' = \bar{x}$



Best choice must be $\alpha = 0$, else we get decrease in f.

Now,

$$Df(\bar{x})^T d_j = D_{d_j} f(\bar{x}) = 0.$$

Since true for a basis of vectors $d_1, \ldots, d_n$, must have

$$Df(\bar{x}) = 0.$$

$\blacksquare$

Notes:

1. No closedness assumptions made on set D.

2. Can be extended to ~~either~~ variable set of directions D_k ~~satisfying~~ ~~either~~ provided
~~(1)~~ they converge to a lin indep set D.
~~or (2) they are~~

3. Need uniqueness along each direction. Else, eg:

$$f(x) = x_2(1-x_1), \quad d_1 = (1,0)^T, \; d_2 = (0,1)^T.$$

Starting at $(0,0)$, ~~get~~ could take any λ_1 in direction d_1,
so eg, $\lambda_1 = 1$, giving $x = (1,0)$.

Then minimizing in direction d_2 allows any λ_2.

So could finish at any point $(1, x_2)$ (or at any point $(x_1, 0)$)

But $\nabla f(x) = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix} = 0$ only at $(0,0)$.