<u>Fixed point algorithms for solving $h(x)=0$</u>

where : $h : \mathbb{R}^n \rightarrow \mathbb{R}^n$

e.g. for solving optimality condition $\nabla V(x)=0$

potentially relevant alg. when
$V \neq$ quadratic and/or n = very big.

<u>Idea</u> : Rewrite $h(x)=0$ as $x = f(x)$

if possible

Then $(\hat{x} = \text{a solution}) \Leftrightarrow [h(\hat{x})=0]$

$\Leftrightarrow [\hat{x} = f(\hat{x})]$

so if $\hat{x}$ goes into f, then $\hat{x}$ comes out, i.e. $\hat{x}$ is a <u>fixed-point</u> of f

Suggests : "fixed-point" alg. :

$$x_{j+1} = f(x_j) \quad ; \quad x_0 \in \mathbb{R}^n \ , \forall j$$

Convergence of x_j to $\hat{x}$?
Stopping condition ?

Analysis needs :

<u>Definition</u> : f is called γ - Lipschitz (on $\mathbb{R}^n$)
if :

$$\|f(x)-f(y)\| \leq \gamma \|x-y\| \ , \forall x,y \in \mathbb{R}^n$$

$\boxed{\gamma \geq 0}$

smallest possible γ will be best for us

If $\gamma < 1 \Rightarrow f$ is called a contraction

Eg. if $f(x) = a + Bx$, $a \in \mathbb{R}^n$, $B \in \mathbb{R}^{n \times n}$

$$\Rightarrow \|f(x)-f(y)\| = \|(a+Bx)-(a+By)\|$$
$$\leq \|B(x-y)\|$$
$$\leq \|B\| \, \|x-y\| \ , \forall x,y \in \mathbb{R}^n$$

$\Rightarrow$ can use $\gamma = \|B\|$, where :

$\|B\| = \|B\|_2$ = induced norm $\triangleq \sqrt{\lambda_{max}(B^T B)}$,
or
$\|B\| = \|B\|_F$ = Frobenius n. $\triangleq \sqrt{\sum_i \sum_j b_{ij}^2}$,

$\|B\|_F \geq \|B\|_2$ (so $\|B\|_F$ not best for γ)
BUT $\|B\|_F$ is much easier to compute

(10.1) **Th** If f is a contraction then:

$$x = f(x)$$

has a unique solution $\hat{x}$.

Proof : (only uniqueness, existence harder)

$$\text{non-unique solution}$$
$$\Downarrow$$
$$\exists \text{ solutions } \bar{x} \text{ and } \tilde{x} \text{ with } \bar{x} \neq \tilde{x}$$
$$\Downarrow$$

$$\begin{aligned} \|\bar{x} - \tilde{x}\| &= \|f(\bar{x}) - f(\tilde{x})\| \\ &\leq \underbrace{\gamma}_{\boxed{<1}} \|\bar{x} - \tilde{x}\| \\ &< \|\bar{x} - \tilde{x}\| \quad \leftarrow \boxed{\text{contradiction}} \end{aligned}$$

Hence : unique solution.

QED

(10.2) **Th** Consider : $x_{j+1} = f(x_j)$; $x_0 \in \mathbb{R}^n$, $\forall j$.

If f is γ-Lipschitz with $\gamma < 1$

(i.e. a contraction)

then:

$\exists$ a unique solution $\hat{x}$ of $x = f(x)$, and:

(i) $0 \leq \|x_j - \hat{x}\| \leq \gamma^j \|x_0 - \hat{x}\|$, $\forall j \geq 0$

so, since $\gamma < 1$

(ii) $x_j \to \hat{x}$ $\leftarrow$ $\boxed{\text{from (i): smaller } \gamma \text{ implies faster, guaranteed convergence}}$

Proof of (i):

$$\begin{aligned} \|x_{j+1} - \hat{x}\| &= \|f(x_j) - f(\hat{x})\| \\ &\leq \gamma \|x_j - \hat{x}\| \end{aligned} \qquad (10.3)$$

so, iterating this,

$$\|x_j - \hat{x}\| \leq \gamma^j \|x_0 - \hat{x}\| , \forall j \geq 0$$

$\Rightarrow$ (i)

QED

A stopping condition

Suppose we would be satisfied with an x_j near enough to $\hat{x}$ in that:

$$\|x_j - \hat{x}\| < \varepsilon \leftarrow \boxed{\text{pre-specified e.g. } 10^{-8}}$$

Not directly useful as stopping-condition since $\hat{x}$ = unknown.

However:

(10.4) **Th** If f is γ-Lipschitz with $\gamma < 1$, then:

$$\|x_j - \hat{x}\| < \frac{\gamma}{1-\gamma} \|x_j - x_{j-1}\| \quad , \forall j > 1$$

> so we can stop iterating, knowing that
> $$\|x_j - \hat{x}\| < \varepsilon$$
> if we continue iterating until
> $$\underbrace{\frac{\gamma}{1-\gamma} \|x_j - x_{j-1}\| < \varepsilon}_{\boxed{\text{easy to test}}}$$

Proof of Th (10.4):

i.e. of $\|x_j - \hat{x}\| \leq \frac{\gamma}{1-\gamma} \|x_j - x_{j-1}\|$ if $\gamma < 1$

Now:

$$\|x_{j-1} - \hat{x}\| = \|x_{j-1} - x_j + x_j - \hat{x}\|$$
$$\leq \|x_{j-1} - x_j\| + \underbrace{\|x_j - \hat{x}\|}_{\boxed{(10.3) \Rightarrow \leq \gamma \|x_{j-1} - \hat{x}\|}}$$
$$\leq \|x_{j-1} - x_j\| + \gamma \|x_{j-1} - \hat{x}\|$$

So: $$\underbrace{(1-\gamma)}_{\boxed{>0 \text{ since } \gamma < 1}} \|x_{j-1} - \hat{x}\| \leq \|x_{j-1} - x_j\|$$

so $$\|x_{j-1} - \hat{x}\| \leq \frac{1}{1-\gamma} \|x_{j-1} - x_j\|$$

Finally

$$\|x_j - \hat{x}\| \underset{\boxed{(10.3)}}{\leq} \gamma \|x_{j-1} - \hat{x}\|$$
$$\leq \frac{\gamma}{1-\gamma} \|x_{j-1} - x_j\|$$
$$= \frac{\gamma}{1-\gamma} \|x_j - x_{j-1}\| \qquad \text{QED}$$

The fixed-point alg. can now be coded as:

- $x_{new} :=$ initial estimate x_0 of $\hat{x}$

Repeat:

- $x_{old} := x_{new}$
- $x_{new} := f(x_{new})$

Until : $\frac{\gamma}{1-\gamma} \| x_{new} - x_{old} \| < \varepsilon$

Question: does it ever terminate?
Answer: yes - always if only $\gamma < 1$.

About rewriting $h(x)=0$ as $x = f(x)$ for the linear case

$$0 = h(x) \triangleq Ax - b \quad , A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n$$

Methodology:

split A as $A = \underbrace{K}_{\boxed{\text{easily invertible approximation to } A}} + \underbrace{E}_{\boxed{\text{error}}}$

Then:
$$0 = Ax - b$$
$$= Kx + Ex - b$$
$$\Rightarrow \quad Kx = b - Ex$$
$$\Rightarrow \quad x = K^{-1}[b - Ex]$$

$\boxed{\text{the re-written } h(x)=0}$

So:
$$\left\{ \begin{array}{l} \text{the split} \\ A = K + E \end{array} \right\} \Rightarrow \quad x = K^{-1}[b - Ex] \triangleq f(x)$$
$$\Downarrow$$
$$f = \underbrace{\| K^{-1} E \|}_{\gamma} \text{ - Lipschitz}$$

Different choices for K

$\Downarrow$

different $f(x)$'s and different fixed-point algs.

$$x_{j+1} = f(x_j)$$

$\Downarrow$

$\| x_j - \hat{x} \| \le \gamma^j \| x_0 - \hat{x} \|$ for different γ's

$\Downarrow$

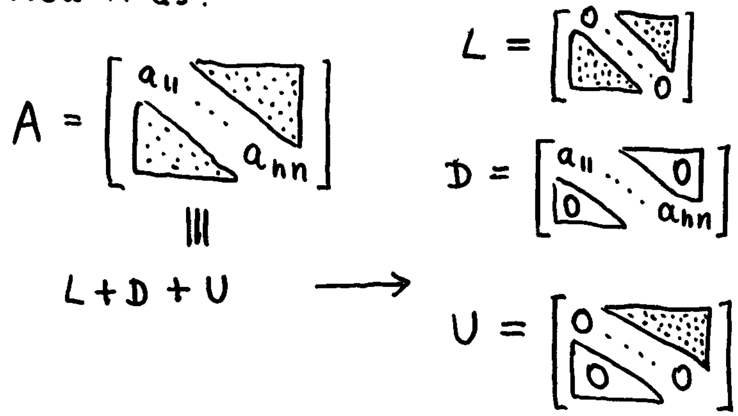
try to choose easily invertible K, so γ = small (est)

best choice depends on A

<u>Some famous choices for K</u>

<u>for the important case when : $a_{ii} \neq 0, \forall i$</u>

View A as :



We want : $A = \underbrace{K}_{\boxed{\text{easily invertible approx. to } A}} + E$

<u>Jacobi iteration :</u>

choose $\underbrace{K = D}_{\boxed{\text{diagonal = easily invertible}}}$, $E = L + U$

<u>Gauss-Seidel iteration :</u>

choose : $\underbrace{K = L + D}_{\boxed{\text{lower-triangular = easily invertible}}}$, $E = U$

<u>Relaxation iteration :</u>

choose : $K = K_\omega \triangleq L + \frac{D}{\omega}$ $\boxed{\text{scalar}}$

$$\Rightarrow \quad E = E_\omega \triangleq -\frac{D}{\omega} + D + U$$

giving : $\gamma_\omega = \| K_\omega^{-1} E_\omega \|$

and try to choose ω so γ_ω = small

ω = so-called relaxation factor

$\omega > 1$: over-relaxation

$\omega < 1$: under-relaxation

$\omega = 1$: relaxation = Gauss-Seidel

Final remarks on fixed-point algs.:

- For a given A, sometimes it might not be possible to find a K so that

  $\|K^{-1}E\| = \gamma$ is less than one.

  However, a suitable K exists sufficiently often for fixed-point algorithms to be of interest.

- fixed-point algs. converge to $\hat{x}$ if only $f = \gamma$-Lipschitz and $\gamma < 1$

  may be linear or nonlinear

- fixed-pt. algs. = simple algs.
- require little storage

  } — GOOD

BUT

- fixed-pt. algs. fail (almost always) if $\gamma > 1$