Fixed Point Iteration

4:51 PM

Finding fixed point of operator F, $F(x)=x$

5.1. Contraction mapping algorithm: [Convergence of contraction mapping iteration]

with contraction factor (Lipschitz constant) 0<L<1

$\{F\{:\mathbb{R}^n\to\mathbb{R}^n, \text{contraction}\}\}$ $x^{k+1}=F(x^k)$ {contraction mapping algorithm} For contraction mapping F, x^{k+1}=F(x^k) converges

Proof: # Strategy: $P\to(Q\wedge R) \equiv (P\to Q)\wedge(Q\to R)$ prove each seperately $\Rightarrow \left(\lim_{k\to\infty} x^k = x^* \;\wedge\; x^* \text{ is unique a fixed point}\right)$

(Uniqueness)

Assume, $x^*, \tilde{x}$ two fixed points of F, $F(\tilde{x})=\tilde{x}$, $F(x^*)=x^*$, $x^*\neq\tilde{x}$

$$\|x^*-\tilde{x}\| = \|F(x^*)-F(\tilde{x})\| \leq \underset{<1}{L}\,\|x^*-\tilde{x}\|$$

$$\to \quad \|x^*-\tilde{x}\| \leq \{<1\}\,\|x^*-\tilde{x}\|$$

$$\leftrightarrow \quad 1 \leq \{<1\} \Rightarrow \text{contradiction} \quad \therefore x^* \text{ unique} \;\blacksquare$$

[eq: uniqueness of contraction mapping fixed point]

(Convergence) assume $L\geq 1$

$$\|x^{k+L}-x^k\| = \|x^{k+L}-x^{k+L-1}+x^{k+L-1}-x^{k+L-2}\ldots-x^{k+1}+x^{k+1}-x^k\|$$

$$= \|(x^{k+L}-x^{k+L-1})+(x^{k+L-1}-x^{k+L-2})+\ldots+(x^{k+1}-x^k)\|$$

$$\leq \|x^{k+L}-x^{k+L-1}\|+\|x^{k+L-1}-x^{k+L-2}\|+\ldots+\|x^{k+1}-x^k\| \quad \{\text{triangle inequality}\}$$

$\|x^{k+2}-x^{k+1}\| = \|F(x^{k+1})-F(x^k)\|_2 \leq L\|x^{k+1}-x^k\|$

$\|x^{k+3}-x^{k+2}\| = \|F(x^{k+2})-F(x^{k+1})\| \leq L\|x^{k+2}-x^{k+1}\| \leq L^2\|x^{k+1}-x^k\|$

Similarly, $\vdots$

$\|x^{k+L}-x^{k+L-1}\| \leq L^{(L-1)}\|x^{k+1}-x^k\|$

$$\leq (L^{L-1}+L^{L-2}+\ldots+L^0)\,\|x^{k+1}-x^k\|$$

$$= \frac{(1-L^L)}{(1-L)}\|x^{k+1}-x^k\|$$

{ $L<1 \Rightarrow (1-L^L) <<1$ so [illegible] }

$$\leq \frac{1}{1-L}\|x^{k+1}-x^k\|$$

[illegible] $\|x^{k+1}-x^k\| = \|F(x^k)-F(x^{k-1})\| \leq L\|x^k-x^{k-1}\| \leq L^2\|x^{k-1}-x^{k-2}\| \leq L^{k-1}\|x^2-x^{k-(k-1)}\| = L^{k-1}\|x^2-x^1\|$

$\|x^k-x^{k-1}\| = \|F(x^{k-1})-F(x^{k-2})\| \leq L\|x^{k-1}-x^{k-2}\|$

$\vdots$

$\|x^3-x^2\| \leq L\|x^2-x^1\|$ }

$$\leq \frac{L^{k-1}}{1-L}\|x^2-x^1\|$$

$$\therefore \|x^{k+L}-x^k\| \leq \frac{L^{k-1}}{1-L}\|x^2-x^1\|$$

$$\forall_{L\geq 1} \lim_{k\to\infty} \underbrace{0\leq\|x^{k+L}-x^k\|}_{\text{positivity of norm}} \leq 0 \quad \left\{\because \lim_{k\to\infty} \underbrace{\frac{1}{1-L}}_{\text{constant}>1} L^{k-1} = 0\right\}$$

$$\therefore \forall_{L\geq 1} \lim_{k\to\infty}\|x^{k+L}-x^k\|=0$$

$$\therefore \forall_{L\geq 1} \lim_{k\to\infty} x^{k+L}=\ldots=x^{k+1}=x^k=x^*$$

now $\{x^i\}$ converges to some constant point x^*, now we want to show $x^*=F(x^*)$ = fixed point of $F(\cdot)$

$\|F(x^*)-x^*\| = \|F(x^*)-x^{k+1}+x^{k+1}-x^*\|$ # Our goal is to show that ||F(x*)-x*|| goes to zero which proves x* is a fixed point

$$= \|(x^{k+1}-x^*)+(F(x^*)-x^{k+1})\|$$

$$\leq \|x^{k+1}-x^*\| + \|(F(x^*)-x^{k+1})\| \quad \#\text{triangle inequality}$$

$$= \|x^{k+1}-x^*\| + \|\underbrace{x^{k+1}}_{F(x^k)}-F(x^*)\| \quad \# \because \|\alpha x\| = |\alpha|\,\|x\|, \; \alpha=-1$$

$$\leq \|x^{k+1}-x^*\| + L\|x^k-x^*\|$$

$$\therefore \forall_k \quad 0\leq\|F(x^*)-x^*\| \leq \|x^{k+1}-x^*\| + L\|x^k-x^*\|$$

$$\lim_{k\to\infty} \quad 0\leq\|F(x^*)-x^*\| \leq \underbrace{\lim_{k\to\infty}\|x^{k+1}-x^*\|}_{0} + \underbrace{\lim_{k\to\infty} L\|x^k-x^*\|}_{0}$$

$$\therefore \|F(x^*)-x^*\|=0 \leftrightarrow F(x^*)=x^* \leftrightarrow x^* \text{ is a fixed point of } F(\cdot)$$

// already proven that such fixed point will be unique [eq: uniqueness of contraction mapping fixed point]

$\blacksquare$

• Another important property:

The distance to the fixed point decreases at each step: (Such algorithm is called Fejér monotone)

[eq: Fejer monotone]

$$\|x^{k+1}-x^*\| = \|F(x^k)-F(x^*)\| \leq L\|x^k-x^*\|$$

$$\leftrightarrow 0\leq \frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|} \leq L\in(0,1) \quad \therefore (\cdot)$$

also using same logic: $\|x^k-x^*\| \leq L\|\underbrace{x^{k-1}}_{F(x^{k-2})}-\underbrace{x^*}_{F(x^*)}\| \leq L^2\|x^{k-2}-x^*\| \ldots \leq L^k\|x^{0=(k-k)}-x^*\|$

$\therefore$ So convergence is atleast geometric with factor L.

• Contraction Mapping algorithm is Used in huge number of applications

• Approach is set up the problem you want to solve as a fixed point equation with a contractive operator.

hard part: showing the operator is contractive!

*** Gradient Method:**

* For a differentiable convex function f: $\quad x^*=\underset{x}{\operatorname{argmin}} f(x) \leftrightarrow \nabla f(x^*)=0 \Leftrightarrow \underbrace{(I+\nabla f)^{-1}}_{\text{proximal operator}}x=x$

$$\underbrace{\{f\{:\mathbb{R}^n\to\mathbb{R}, \delta_{\mathbb{I}}, \mathcal{D}\}, x_0\{\text{starting point}, \in\mathbb{R}^n, \alpha\{>0, \text{stepsize}\}\}\}}_{\cdot}\; \{\text{gradient method}\}\; x^{k+1}=x^k-\alpha\nabla f(x^k) = \underbrace{(I\cdot-\alpha\nabla f(\cdot))}_{G}(x^k)$$

from this:

$$x^*=G(x^*) \leftrightarrow x^*=\underset{x}{\operatorname{argmin}} f(x)$$

$$(f\{\cdot, \text{strongly convex}, \text{strongly smooth}) \Rightarrow G=(I\cdot-\alpha\nabla f(\cdot))\{\text{Lipschitz with parameter } L_{G,M}=\max\{|1-\alpha m|, |1-\alpha L|\}\}$$

$\exists_{m}(f(x)-m\|x\|_2^2)\{\mathcal{D}\}$ / $\exists_{L}(f(x)-L\|x\|_2^2)\{\mathcal{D}\}$

The intuition behind strongly convex and strongly smooth is that if we take a small amount of norm squared convexity the fu nction still stays convex, but if there is point of how much we can take away this norm squared convexity. At this breaking point, the fun ction with so much convexity taken away flips its nature and becomes concave. Clearly L>m

$\left(f\{\Box\},\ \text{strongly convex},\ \text{strongly smooth} \right) \Rightarrow G = (I\Box - \alpha\nabla f(\Box))\ \{\text{Lipschitz with parameter } L_{G,M} = \max\{|1-\alpha m|, |1-\alpha L|\}\}$

$\exists_{m>0} \left(f(x) - \frac{m}{2}\|x\|_2^2 \right) \{D\}$ $\exists_{L>0} \left(f(x) - \frac{L}{2}\|x\|_2^2 \right) \{\cap\}$

The intuition behind strongly convex and strongly smooth is that if we take a small amount of norm squared convexity the fu nction still stays convex, but if there is point of how much we can take away this norm squared convexity. At this breaking point, the fun ction with so much convexity taken away flips its nature and becomes concave. Clearly L>m

• Proof: [# Required information:

- $F\Box_{\{:\mathbb{R}^n\to\mathbb{R}^n,\ \delta_e\}}$ {Lipschitz with parameter L} $\Leftrightarrow \forall_x\ \|DF(x)\|_2 \leq L$ (means $(D \circ F)(x)$)

- $(D\circ\nabla)(\Box) = \nabla^2(\Box)$

  i.e. $(D\circ\nabla)(f(x)) = (D\nabla)(f(x)) = \nabla^2 f(x)$ #]

So, [illegible] $\forall_x\ \|DG(x)\|_2 \leq \tilde{L}$ then $\tilde{L}$ will be the Lipschitz constant for relation F

$(D\circ\nabla)(f(x)) = (\nabla^2)(f(x)) = \nabla^2 f(x)$

$DG(x) = (D\circ G)(x) = D(G(x)) = D(x - \alpha\nabla f(x)) = Dx - \alpha\left(D\circ(\nabla f(x))\right) = I - \alpha\nabla^2 f(x) \quad \ldots \quad (1)$

$(I\Box - \alpha\nabla f(\Box))(x) = (\underset{x}{\underset{\|}{Ix}} - \alpha\nabla f(x))$

$Dx = \left[\left[\frac{\partial}{\partial x_j} x_i\right]_{i=1}^{n}\right]_{j=1}^{n} = ((\delta_{ij})_i)_j = I$ (row, col)

now f is strongly convex $\Leftrightarrow \forall_x \left(f(x) - \frac{m}{2}\|x\|_2^2\right)\{D\}$ $\{\delta_e\}$

$\Leftrightarrow \nabla^2\left(f(x) - \frac{m}{2}\|x\|_2^2\right) \succeq 0$

$\Leftrightarrow \nabla^2 f(x) - \frac{m}{2}\nabla^2\|x\|_2^2 = \nabla^2 f(x) - \frac{m}{2}\cdot 2\cdot I = \nabla^2 f(x) - mI \succeq 0 \Leftrightarrow \nabla^2 f(x) \succeq mI$

$(D\circ\nabla)(\|x\|_2^2) = D(\underbrace{\nabla(\|x\|_2^2)}_{2x}) = 2D(x) = 2I$

$\Leftrightarrow -\alpha\nabla^2 f(x) \preceq -\alpha m I$

$\Leftrightarrow \underbrace{I - \alpha\nabla^2 f(x)}_{DG(x)} \preceq I - \alpha m I = (1-\alpha m)I$

generalized inequalities satisfy the common properties of normal inequalities (one exception: total ordering [illegible] partial ordering works in generalized inequality)

f is strongly smooth

$\Leftrightarrow \forall_x \left(f(x) - \frac{L}{2}\|x\|_2^2\right)\{\cap\}$

$\Leftrightarrow \forall_x\ \nabla^2\left(f(x) - \frac{L}{2}\|x\|_2^2\right) \preceq 0 \Leftrightarrow \nabla^2 f(x) - LI \preceq 0 \Leftrightarrow \nabla^2 f(x) \preceq LI$

$\nabla^2 f(x) - LI$ #just like before#

$\Leftrightarrow -\alpha\nabla^2 f(x) \succeq -\alpha L I$

$\Leftrightarrow \underbrace{I - \alpha\nabla^2 f(x)}_{DG(x)} \succeq I - \alpha L I = (1-\alpha L)I$

$\Rightarrow \forall_x\ (1-\alpha L)I \preceq DG(x) \preceq (1-\alpha m)I$

$(1-\alpha m)I - DG(x) \succeq 0 \ \wedge\ (DG(x) - (1-\alpha L)I) \succeq 0$

$\forall_y\ y^T\left((1-\alpha m)I - DG(x)\right)y \geq 0 \ \wedge\ y^T\left(DG(x) - (1-\alpha L)I\right)y \geq 0$

$\Leftrightarrow (1-\alpha m)y^Ty - y^TDG(x)y \geq 0 \ \wedge\ y^TDG(x)y \geq y^T(1-\alpha L)y$

$\Leftrightarrow (1-\alpha m)y^Ty \geq y^TDG(x)y \ \wedge\ y^TDG(x)y \geq (1-\alpha L)y^Ty$

$\Leftrightarrow (1-\alpha L)\|y\|_2^2 \leq y^TDG(x)y \leq (1-\alpha m)\|y\|_2^2$

$\Leftrightarrow \forall_{y\neq 0}\ (1-\alpha L) \leq \frac{y^TDG(x)y}{\|y\|_2^2} \leq (1-\alpha m)$

$y := v_{\lambda_1} \Rightarrow (1-\alpha L) \leq \lambda_1 \leq (1-\alpha m)$

$\therefore |\lambda_1| \leq \max\{|1-\alpha m|, |1-\alpha L|\}$

$x \leq y \leq z \Rightarrow |y| \leq \max\{|x|, |z|\}$

(check it for different x, y, z, it is true e.g.,

$-1 \leq 2 \leq 3 \Rightarrow |2| \leq \max\{|-1|, |3|\} = 3$

$-3 \leq -2 \leq -1 \Rightarrow |-2| \leq \max\{|-1|, |-3|\} = 3$ #)

for any matrix, $\|A\|_2 = \max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2}$

if A is symmetric,

now $A \in S^n \Leftrightarrow A = Q\Lambda Q^T$, $A = Q\Lambda^2 Q^T$

$\|Ax\|_2^2 = x^TA^TAx = x^TA^TAx = x^TQ\Lambda^2Q^Tx = \underbrace{(Q^Tx)}_{y}\Lambda^2\underbrace{(Q^Tx)}_{y} = y\ \mathrm{diag}(\lambda_1^2, \ldots, \lambda_n^2)\,y = [y_1, \ldots, y_n]\begin{bmatrix}\lambda_1^2 y_1\\ \vdots \\ \lambda_n^2 y\end{bmatrix} = \sum_{i=1}^n \lambda_i^2 y_i^2 \leq \lambda_1^2\sum_{i=1}^n y_i^2 = \lambda_1^2\|y\|_2^2 = \lambda_1^2\|Q^Tx\|_2^2 = \lambda_1^2\|x\|_2^2$

$\lambda_1 \geq \ldots \geq \lambda_n$

$\rightarrow \frac{\|Ax\|_2^2}{\|x\|_2^2} \leq \lambda_1^2 \rightarrow \frac{\|Ax\|_2}{\|x\|_2} \leq |\lambda_1|$

for. $x = v_{\lambda_1}$, $\frac{\|Av_{\lambda_1}\|_2^2}{\|v_{\lambda_1}\|_2^2} = |\lambda_1|\frac{\|v_{\lambda_1}\|_2^2}{\|v_{\lambda_1}\|_2^2} = |\lambda_1|$

$\therefore \|A\|_2\ \{A\in S^n\} = |\lambda_1|$

$\|DG(x)\|_2 \leq \max\{|1-\alpha m|, |1-\alpha L|\} \Rightarrow G_{\{:\mathbb{R}^n\to\mathbb{R}^n,\ \delta_e\}}$ {Lipschitz with constant $\tilde{L} = \max\{|1-\alpha m|, |1-\alpha L|\}$} [Proved] ☺

So, if $\tilde{L} = \max\{|1-\alpha m|, |1-\alpha L|\} < 1$ then we have a contraction mapping.

$\Leftrightarrow |1-\alpha m| < 1 \ \wedge\ |1-\alpha L| < 1$

$\Leftrightarrow \max\{1-\alpha m, -1+\alpha m\} < 1 \ \wedge\ \max\{1-\alpha L, -1+\alpha L\} < 1$

$\Leftrightarrow \underbrace{1-\alpha m < 1}_{\alpha > 0} \ \wedge\ \underbrace{-1+\alpha m < 1}_{\alpha < \frac{2}{M}} \ \wedge\ \underbrace{1-\alpha L < 1}_{\alpha > 0} \ \wedge\ \underbrace{-1+\alpha L < 1}_{\alpha < \frac{2}{L}}$

$\Leftrightarrow 0 < \alpha < \min\left\{\frac{2}{L}, \frac{2}{M}\right\} = \frac{2}{L}$

$L > m > 0 \Leftrightarrow \frac{1}{L} < \frac{1}{m}$, so $\min\left\{\frac{2}{L}, \frac{2}{m}\right\} = \frac{2}{L}$

$\therefore 0 < \alpha < \frac{2}{L}$ stepsize selection ensures that G: contraction.

and the $x^{k+1} = G(x^k)$ will converge to unique point x^* #note that we are not showing that x^* is the minimizer, for that we would require other methods.

$G = (I\Box - \alpha\nabla f(\Box))$

Page 14, 15 5.2. Fixed point set of a nonexpansive operator

$F(x) = x,\ F(y) = y,\ \theta\in[0,1],\ z = \theta(x) + (1-\theta)y \overset{?}{\in} X = \{\Box \mid F\Box = \Box\}$

$F(z) = w$

$\|w - x\|_2 = \|F(z) - F(x)\|_2 \leq \|z - x\|_2 = \|\theta x + (1-\theta)y - x\|_2 = \|(1-\theta)y - (1-\theta)x\|_2 = (1-\theta)\|y - x\|_2$

$\|w - y\|_2 = \|F(z) - F(y)\|_2 \leq \|z - y\|_2 = \|\theta x + \underbrace{(1-\theta)y - y}_{y - \theta y - y}\|_2 = \|-\theta(y - x)\|_2 = \theta\|y - x\|_2$

$\Rightarrow \|w - x\|_2 + \|w - y\|_2 \leq \|x - y\|_2$

Again by triangle inequality,

$\|w - x\|_2 + \|w - y\|_2 \geq \|x - y\|_2$

$\Rightarrow \|x - y\|_2 = \|w - x\|_2 + \|w - y\|_2$

triangle inequality equality [illegible] w lies in between on the line segment constructed by x, y.

y —— w —— x

so $\exists_{\bar\theta\in[0,1]}\ w = \bar\theta x + (1-\bar\theta)y$,

$\|w - y\|_2 \leq \theta\|y - x\|_2 \Rightarrow \bar\theta\|y - x\|_2 \leq \theta\|y - x\|_2$

$\|\bar\theta x + \underbrace{(1-\bar\theta)y - y}_{y - \bar\theta y - y}\|_2$

$= \|\bar\theta x - \bar\theta y\|_2 = |\bar\theta|\,\|y - x\|_2$

$\Rightarrow \bar\theta \leq \theta$ $\{x \neq y\}$

assume $\bar\theta < \theta$

$\rightarrow -\bar\theta > -\theta$

$\rightarrow (1-\bar\theta) > (1-\theta)$

but $(1-\bar\theta)\|y - x\|_2 \leq (1-\theta)\|y - x\|_2$

$\rightarrow (1-\bar\theta) \leq (1-\theta)$

$\rightarrow$ contradiction $\Rightarrow \bar\theta = \theta$ $\therefore w = \theta x + (1-\theta)y = z \in X = \{\Box \mid F\Box = \Box\}$

$\therefore X_{\{D\}}$

some contraction mapping

☠: The contraction mapping algorithm $x^{k+1} = \Box(x^k)$, may not work for a nonexpansive mapping! We need damped iteration for them.

* 5.3. Damped iteration for nonexpansive mapping: (from the slides of Boyd)

$\left(x^{k+1} = (1-\theta)x^k + \theta F(x^k) \quad //\text{damped version} \right) \Rightarrow \left(\|F(x^k) - x^k\|_2 \overset{?}{\to} 0 \ \wedge\ \min_{j\in\{0,1,\ldots,k\}} \|F(x^j) - x^j\|_2^2 = O\left(\tfrac{1}{k}\right) \right)$

$\theta \in (0,1)$

this is what we care about as

[# damped iteration for finding fixed point of nonexpansive mapping #]

* 5.3. Damped iteration for nonexpansive mapping:

$$\left(x^{k+1} = (1-\theta)x^k + \theta F(x^k) \quad \text{//damped version} \right) \Rightarrow \left(\begin{array}{l} \|F(x^k) - x^k\|_2 \overset{?}{\to} 0 \ \wedge \\ \min_{j\in\{0,1,\dots,k\}} \|F(x^j) - x^j\|_2^2 = O\left(\frac{1}{k}\right) \end{array} \right)$$

[# damped iteration for finding fixed point of nonexpansive mapping #]

$\theta \in (0,1)$

this is what we care about as stopping criterion will be $\|F(x^k) - x^k\|_2 \leq \epsilon$

$$x^{k+1} = \underbrace{\left((1-\theta) I \boxminus + \theta \underbrace{F(\boxminus)}_{\text{original nonexpansive operator}}\right)}_{\text{damped version of the operator}} (x^k)$$

Conclusion:

$\mathrm{dist}(x^k, X) \overset{?}{\to} 0$ } $\{x^k\}$ is Fejér monotone and the iterates converge to a solution.

Here X = set of all fixed points

x^* = one fixed point

* 5.4. Convergence Proof for damped iteration

identity used: $\forall_{\theta\in\mathbb{R},\, a, b \in \mathbb{R}^n} \ \|\theta a + (1-\theta) b\|_2^2 = \theta\|a\|_2^2 + (1-\theta)\|b\|_2^2 - \theta(1-\theta)\|b-a\|_2^2$ /* Corollary 2.14. Bauschke */

L.H.S $= (\theta a^T + (1-\theta) b^T)(\theta a + (1-\theta) b) = \theta^2\|a\|_2^2 + (1-\theta)^2\|b\|_2^2 + 2\theta(1-\theta) a^T b$

R.H.S $= \theta\|a\|_2^2 + (1-\theta)\|b\|_2^2 - \theta(1-\theta)\underbrace{\|b-a\|_2^2}$

$(b^T - a^T)(b-a) = \|b\|_2^2 + \|a\|_2^2 - 2a^T b$

$-\theta(1-\theta)(\|b\|_2^2 + \|a\|_2^2 - 2a^T b)$

$= \|a\|_2^2 \underbrace{(\theta - \theta(1-\theta))}_{\cancel{\theta} - \cancel{\theta} + \theta^2} + \|b\|_2^2 \underbrace{((1-\theta) - \theta(1-\theta))}_{(1-\theta)^2} + 2\theta(1-\theta) a^T b$

$= \theta^2\|a\|_2^2 + (1-\theta)^2\|b\|_2^2 + 2\theta(1-\theta)a^T b$ = L.H.S #

The damped iteration for non-expansive mapping is:

$$x^{k+1} = \theta x^k + (1-\theta) F(x^k) \quad \{\theta \in (0,1)\}$$

[eq: damped iteration for nonexpansive mapping]

So,

$$\|x^{k+1} - x^*\|_2^2 = \|\theta x^k + (1-\theta) F(x^k) - x^*\|_2^2 = \|\theta\underbrace{(x^k - x^*)}_{a} + (1-\theta)\underbrace{(F(x^k) - x^*)}_{b}\|_2^2 = \theta\|x^k - x^*\|_2^2 + (1-\theta)\|F(x^k) - x^*\|_2^2 - \theta(1-\theta)\underbrace{\|F(x^k) - \cancel{x^*} - x^k + \cancel{x^*}\|_2^2}_{\|F(x^k) - x^k\|_2^2}$$

$F(x^*) = x^*$

$\theta x^k + (1-\theta) F(x^k) - x^*$

$= \theta x^k - \theta x^* + \underbrace{\theta x^* - x^*}_{-x^*(1-\theta)} + (1-\theta) F(x^k)$

$= \theta(x^k - x^*) + (1-\theta)(F(x^k) - x^*)$

$\|F(x^k) - F(x^*)\|_2^2 \leq \underset{\substack{\| \\ 1}}{L}\|x^k - x^*\|_2^2 = \|x^k - x^*\|_2^2$

$\leq \theta\|x^k - x^*\|_2^2 + (1-\theta)\|x^k - x^*\|_2^2 - \theta(1-\theta)\|F(x^k) - x^k\|_2^2$

$= \|x^k - x^*\|_2^2 - \underbrace{\theta(1-\theta)\|F(x^k) - x^k\|_2^2}$

negative [illegible] when $F(x^k) \neq x^k$ i.e. for normal iteration, so [illegible] strictly larger number [illegible]

$\leq \|x^k - x^*\|_2^2$ [equality only for $x^k = x^*$].

$\forall_k \ \|x^{k+1} - x^*\|_2^2 \leq \|x^k - x^*\|_2^2$ [equality only for $x^{k+1} = x^k \leftrightarrow F(x^k) = x^k$ = some fixed point].

eq: Fejer Monotonicity of damped iteration

So for normal iterations $\|x^{k+1} - x^*\|_2^2 < \|x^k - x^*\|_2^2$ so, distance to fixed point every normal iteration [illegible], so this is Fejer monotone !

$\mathrm{dist}(x^k, X) \downarrow$

Note that dist(x^k,X) decreases is not enough, amader dekhate hobe dist(x^k,X) -> 0, as it might reduce to some nonzero value i.e., dist(x^k,X) ->10 meaning no matter what you do x^k cannot reach the fixed point set!

, i.e., suppose $x^* = 1$, then at some point $x^{k+1} = x^k = 10$. then $\|x^{k+1} - x^*\| \leq \|x^k - x^*\| \leftrightarrow \|10-1\|^2 = \|10-1\|^2$ yet it has not reached the fixed point.

$$\|x^{k+1} - x^*\|_2^2 \leq \|x^k - x^*\|_2^2 - \theta(1-\theta)\|F(x^k) - x^k\|_2^2 \leq \|x^{k-1} - x^*\|_2^2 - \left(\theta(1-\theta)\sum_{j=k-1}^{k}\|F(x^j) - x^j\|_2^2\right) \leq \|x^0 - x^*\|_2^2 - \theta(1-\theta)\sum_{j=0}^{k}\|F(x^j) - x^j\|_2^2$$

$\|x^k - x^*\|_2^2 \leq \|x^{k-1} - x^*\|_2^2 - \theta(1-\theta)\|F(x^{k-1}) - x^{k-1}\|_2^2$

$\vdots$

$\|x^1 - x^*\|_2^2 \leq \|x^0 - x^*\|_2^2 - \theta(1-\theta)\|F(x^0) - x^0\|_2^2$

$$0 \leq \|x^{k+1} - x^*\|_2^2 \leq \|x^0 - x^*\|_2^2 - \theta(1-\theta)\sum_{j=0}^{k}\|F(x^j) - x^j\|_2^2$$

$$\rightarrow \theta(1-\theta)\sum_{j=0}^{k}\|F(x^j) - x^j\|_2^2 \leq \|x^0 - x^*\|_2^2$$

$$\rightarrow 0 \leq \sum_{j=0}^{k}\|F(x^j) - x^j\|_2^2 \leq \frac{\|x^0 - x^*\|_2^2}{\theta(1-\theta)}$$

[∵ sum of norms]

constant

From real analysis:

- $\left(s_n = \sum_{k=0}^{n} a_k,\ a_k \geq 0,\ k \geq 0,\ \exists_{M>0}\ \forall_n\ s_n \leq M\right) \Rightarrow s_\infty = \sum_{k=0}^{\infty} a_k$ converges
- $\left(\sum_{k=0}^{\infty} a_k : \text{converges}\right) \Rightarrow \lim_{k\to\infty} a_k = 0$ #

Intuitively it means that if a finite sum of positive terms is globally bounded for any number of terms considered, then infinite sum of those positive terms will converge

$$\left(s_k = \sum_{j=0}^{k}\underbrace{\|F(x_j) - x^j\|_2^2}_{\geq 0 \text{ for } k\geq 0\ ☑},\quad \exists_{M = \frac{\|x^0 - x^*\|_2^2}{\theta(1-\theta)}}\ \forall_k\ s_k \leq M\ ☑\right) \Rightarrow \left(\sum_{j=0}^{\infty} a_n \text{ converges}\right) \Rightarrow \left(\lim_{n\to\infty}\|F(x^j) - x^j\|_2^2 = 0\right)$$

[illegible], $F(x^j) \hookrightarrow x^j$

As: $\sum_{j=0}^{k}\underbrace{\|F(x^j) - x^j\|_2^2}_{\geq 0} \leq \frac{\|x^0 - x^*\|_2^2}{\theta(1-\theta)}$

Now, by defn, $\forall_{j\in\{0,\dots,k\}}$ $\min_{\bar{j}\in\{0,\dots,k\}}\{\|F(x^{\bar{j}}) - x^{\bar{j}}\|_2^2\} \leq \|F(x^j) - x^j\|_2^2$

because this is a fixed term, so [illegible]

$$\rightarrow \sum_{j=0}^{k}\left(\min_{\bar{j}\in\{0,\dots,k\}}\{\|F(x^{\bar{j}}) - x^{\bar{j}}\|_2^2\}\right) \leq \sum_{j=0}^{k}\|F(x^j) - x^j\|_2^2 \leq \frac{\|x^0 - x^*\|_2^2}{\theta(1-\theta)}$$

$$\leftrightarrow \min_{\bar{j}\in\{0,\dots,k\}}\{\|F(x^{\bar{j}}) - x^{\bar{j}}\|_2^2\}\cdot(k+1) \leq \frac{\|x^0 - x^*\|_2^2}{\theta(1-\theta)}$$

$$\therefore \min_{\bar{j}\in\{0,\dots,k\}}\{\|F(x^{\bar{j}}) - x^{\bar{j}}\|_2^2\} \leq \frac{\|x^0 - x^*\|_2^2}{(k+1)\theta(1-\theta)}$$

[Note that this equation only says that the best solution so far will approach as we take more and more iterates, but it does not implies that the iterate improves to get better solutions as we give more iterations, we want to show that $\lim_{k\to\infty} x^k = x^*$

We want (·) as close to zero as possible, but this approaches to zero

$\therefore \min_{j \in \{0,\dots,k\}} \{\|F(x^j) - x^j\|_2^2\} \leqslant \frac{\|x^0 - x^*\|_2^2}{(k+1)\theta(1-\theta)}$

[Note that this equation only says that the best solution so far will approach as we take more and more iterates, but it does not implies that the iterate improves to get better solutions as we give more iterations, we want to show that $\lim_{k\to\infty} x^k = x^*$

We want $(\cdot)$ as close to zero as possible, but this approaches to zero at a linear $O\left(\frac{1}{k}\right)$ rate, we would have liked $O\left(\frac{1}{e^k}\right)$ or something like that.

↳ pretty bad. ☹

Now we want to show: $\exists_{x^* \in X} \operatorname{Lt}_{k\to\infty} x^k = x^*$

is $x \in X$ the $\{x^k\}$ sequence of iterates lie within the compact set $\{z \in \mathbb{R}^n \mid \|z - x\|_2 \leqslant \|x^0 - x\|_2\}$

this is a compact subset in $\mathbb{R}^n$ (which is a metric space) so, the sequence $\{x^k\}$ lies in a compact set

eg: Fejer Monotonicity of damped iteration

$\Rightarrow$ the sequence $\{x^k\}$ must have a limit point

∵ Rudin Theorem 3.11 :

$$\left(\left(X \text{ compact metric space}, \{p_n\} \text{ cauchy sequence in } X\right) \Rightarrow \exists_{p^* \in X} \; p_n \to p^*\right)$$

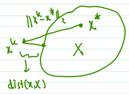
$$\left(\exists_{x^* \in \{z \in \mathbb{R}^n \mid \|z - x\|_2 \leqslant \|x^0 - x\|_2\}} \; x^k \to x^*\right)$$

$$\left(F(x^k) \to x^k \wedge x^k \to x^*\right) \Rightarrow F(x^*) = x^*$$

So, $\left(\{x^k\} \text{ satisfies Fejer monotonicity } \|x^{k+1} - x^*\| \leqslant \|x^k - x^*\| \text{ and } x^k \to x^*\right) \Rightarrow \{x^k\}_0 \to x^*$

equality when $x^k = x^*$

now, $0 \leqslant \operatorname{dist}(x^k, X) \leqslant \|x^k - x^*\|_2$ $\quad \therefore \operatorname{Lt}_{k\to\infty} \; 0 \leqslant \operatorname{Lt}_{k\to\infty} \operatorname{dist}(x^k, X) \leqslant \operatorname{Lt}_{k\to\infty} \|x^k - x^*\|_2 = 0$



$\therefore 0 \leqslant \operatorname{Lt}_{k\to\infty} \operatorname{dist}(x^k, X) \leqslant 0 \Leftrightarrow \operatorname{dist}(x^k, X) \to 0$ ☺

[Second reading]

■ Averaged operators:

An operator F is averaged if it can be written as the convex combination of identity and some nonexpansive operator.

I: identity, G:nonexpansive // (1-θ) I + θ G:θ ∈ (0,1)=Averaged operator

* F: averaged operator,

$\tilde{F}$: averaged operator $\rightarrow$ $F\tilde{F}$: averaged operator

Fixed point finding for averaged operator:

$F: \mathbb{R}^n \to \mathbb{R}^n$, averaged, then to find a fixed point of $F(x) = x$ it suffices to run $x^{k+1} = F(x^k)$ and it will converge to a fixed point if there exists one.

$\exists_{G:\text{nonexpansive}} \; F = \theta I + (1-\theta) G$

Suppose the set of fixed points is nonempty. Then $\exists x^* \in X$ $\quad x^k \to x^*$, also the algorithm will be Fejer monotone, i.e.,

$$\operatorname{dist}(x^k, X) = \inf_{z \in X} \|z - x\|_2^2 \to 0 \text{ : monotonously}$$

Also the rate of convergence:

$$\|Fx^k - x^k\|_2 \to 0 \text{ with rate } \min_{j \in \{0,\dots,k\}} \|Fx^j - x^j\|_2^2 = O\left(\frac{1}{k}\right)$$