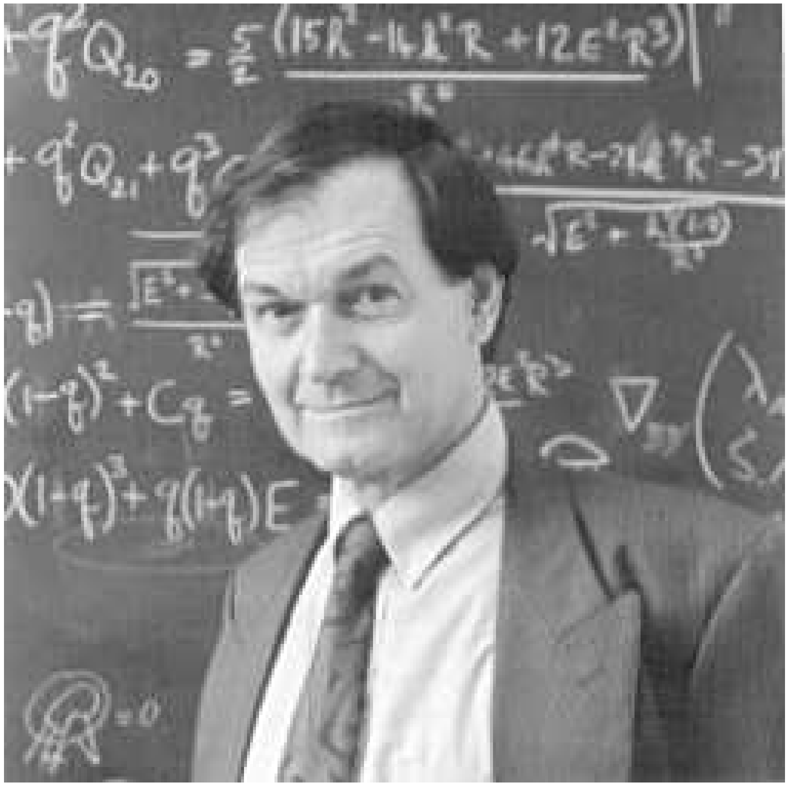
Lecture 2: Generalized Inverses



Moore's plan

The striking analogies between the theories for linear equations in n–dimensional Euclidean space, for Fredholm integral equations in the space of continuous functions defined on a finite real interval, and for linear equations in Hilbert space of infinitely many dimensions, led Moore to lay down his well–known principle.

"The **existence of analogies** between central features of various theories **implies** the **existence** of a more fundamental **general theory** embracing the special theories as particular instances and unifying them as to those central features." (Moore, 1912)

"The **effectiveness** of the **reciprocal** of a non–singular finite matrix in the study of properties of such matrices makes it **desirable** to define if possible **an analogous matrix** to be associated with **each finite matrix** even if it is **not square** or, if square, is **not necessarily non–singular**." (Moore 1935)

Desiderata

$\mathbb{C}_r^{m \times n}$ = the $m \times n$ matrices over $\mathbb{C}$ with **rank** r.

A matrix $A \in \mathbb{C}^{n \times n}$ is **nonsingular** if rank $A = n$, or det $A \neq 0$.
The **inverse** of A satisfies, by definition, the following equations,

$$AXA = A \tag{1}$$

$$XAX = X \tag{2}$$

$$(AX)^* = AX \tag{3}$$

$$(XA)^* = XA \tag{4}$$

$$AX = XA \tag{5}$$

as well as the conditions

$$A\mathbf{x} = \lambda\,\mathbf{x} \implies A^{-1}\mathbf{x} = \frac{1}{\lambda}\,\mathbf{x} \tag{6}$$

$$A, B \text{ nonsingular} \implies (AB)^{-1} = B^{-1}A^{-1} \tag{7}$$

These properties are desirable, can one have them for general A ?

The Penrose equations

The **Penrose equations** for $A \in \mathbb{C}^{m \times n}$ are:

$$AXA = A\,, \tag{1}$$

$$XAX = X\,, \tag{2}$$

$$(AX)^* = AX\,, \tag{3}$$

$$(XA)^* = XA\,. \tag{4}$$

Let $A\{i, j, \ldots, k\}$ denote the set of matrices $X \in \mathbb{C}^{n \times m}$ which satisfy equations $(i), (j), \cdots, (k)$.

A matrix $X \in A\{i, j, \ldots, k\}$ is called an $\{i, j, \ldots, k\}$–**inverse** of A, and also denoted by $A^{(i,j,\ldots,k)}$.

In particular, a $\{1\}$–**inverse**, a $\{2\}$–**inverse**, a $\{1, 3\}$–**inverse**, etc.

The **Moore–Penrose inverse** of A is its $\{1, 2, 3, 4\}$–inverse, denoted $A^\dagger$.

Why Moore's work was unknown in 1955?

Answer: Telegraphic style and idiosyncratic notation. Example:

(29.3) **Theorem**.

$\mathfrak{U}^{C}\,\mathfrak{B}^{1\ \mathrm{II}}\,\mathfrak{B}^{2\ \mathrm{II}}\,\kappa^{12}\cdot)\,\cdot$

$\exists\,|\,\lambda^{21\ \text{type}\ \mathfrak{M}^{2}_{\kappa *}\ \overline{\mathfrak{M}}^{1}_{\kappa}}\ \ni\ \cdot S^{2}\,\kappa^{12}\,\lambda^{21} = \delta^{11}_{\mathfrak{M}^{1}_{\kappa}}\ \cdot\ S^{1}\,\lambda^{21}\,\kappa^{12} = \delta^{22}_{\mathfrak{M}^{2}_{\kappa *}}$

English translation:

(29.3) **Theorem**.

For every matrix A there exists a unique matrix $X : R(A) \to R(A^{*})$ such that

$$AX = P_{R(A)}\ ,\ XA = P_{R(A^{*})}\ .$$

Construction of $\{1\}$–inverses

Given $A \in \mathbb{C}_r^{m\times n}$, let $E \in \mathbb{C}_m^{m\times m}$ and $P \in \mathbb{C}_n^{n\times n}$ be such that

$$EAP = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} . \tag{1}$$

Then for any $L \in \mathbb{C}^{(n-r)\times(m-r)}$, the $n \times m$ matrix

$$X = P \begin{bmatrix} I_r & O \\ O & L \end{bmatrix} E \tag{2}$$

is a $\{1\}$–inverse of A. The partitioned matrices in (1), (2) must be suitably interpreted in case $r = m$ or $r = n$.

Proof. Write (1) as

$$A = E^{-1} \begin{bmatrix} I_r & K \\ O & O \end{bmatrix} P^{-1} ,$$

then verify that any X given by (2) satisfies $AXA = A$. □

Linear equations

Given $A \in \mathbb{C}^{m \times n}$, $\mathbf{b} \in \mathbb{C}^m$, the equations

$$A\,\mathbf{x} = \mathbf{b} \tag{1}$$

have a solution if and only if for any $X \in A\{1\}$,

$$AX\mathbf{b} = \mathbf{b}\ , \tag{2}$$

in which case the general solution is

$$\mathbf{x} = X\,\mathbf{b} + (I - X\,A)\mathbf{y}\ ,\ \ \mathbf{y} \in \mathbb{C}^n \text{ arbitrary} \tag{3}$$

Proof. $AXA = A \implies AX$ idempotent, rank $AX =$ rank A.

$\therefore\ AX = P_{R(A),M}$, for some M such that $C^m = R(A) \oplus M$.

$A\mathbf{x} = \mathbf{b}$ consistent $\iff \mathbf{b} \in \mathbb{R}(A) \iff P_{R(A),M}\mathbf{b} = \mathbf{b}$, $\forall\ M$

Finally, $A\,(X\,\mathbf{b} + (I - X\,A)\mathbf{y}) = AX\,\mathbf{b} = \mathbf{b}$. $\square$

Linear matrix equations

Theorem. Let $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{p\times q}$, $D \in \mathbb{C}^{m\times q}$. Then the matrix equation

$$AXB = D \tag{1}$$

is consistent if and only if for some $A^{(1)}, B^{(1)}$,

$$AA^{(1)}DB^{(1)}B = D \ , \tag{2}$$

in which case the general solution is

$$X = A^{(1)}DB^{(1)} + Y - A^{(1)}AYBB^{(1)} \tag{3}$$

for arbitrary $Y \in \mathbb{C}^{n\times p}$.

Proof. If (1) is consistent then

$$D = AXB = AA^{(1)}AXBB^{(1)}B = AA^{(1)}DB^{(1)}B \ .$$

Kronecker products and matrix equations

The **Kronecker product** $A \otimes B$ of the two matrices $A = (a_{ij}) \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{p\times q}$ is the $mp \times nq$ matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}$$

For $X = (x_{ij}) \in \mathbb{C}^{m\times n}$, let $\mathrm{vec}(X) = (v_k) \in \mathbb{C}^{mn}$ be the vector obtained by listing the elements of X by rows,

$$v_{n(i-1)+j} = x_{ij} \quad (i \in \overline{1,m}\ ;\ j \in \overline{1,n})$$

Lemma. For compatible matrices A, X, B

$$(A \otimes B^T)\,\mathrm{vec}(X) = \mathrm{vec}\,(AXB)$$

Construction of $\{1,2\}$–inverses

Proposition. Let $Y, Z \in A\{1\}$, and let

$$X = YAZ \ .$$

Then $X \in A\{1,2\}$.

Proof. $AXA = A(YAZ)A = (AYA)ZA = AZA = A$,

$$XAX = (YAZ)A(YAZ) = Y(AZA)YAZ = Y(AYA)Z = X \ . \ \square$$

Proposition. Any two of the following statements imply the third:

(a) $X \in A\{1\}$,

(b) $X \in A\{2\}$,

(c) $\operatorname{rank} X = \operatorname{rank} A$.

Proof. $X \in A\{1\},\ Y \in A\{2\} \implies \operatorname{rank} Y \leq \operatorname{rank} A \leq \operatorname{rank} X$, etc.

Projections

Theorem. For any $A \in \mathbb{C}^{m\times n}$, $A^{(1)} \in A\{1\}$.

$$R(AA^{(1)}) = R(A)\ ,\ N(A^{(1)}A) = N(A)\ ,\ R((A^{(1)}A)^*) = R(A^*)\ .$$

Proof. Always $R(AX) \subset R(A)\ ,\ N(A) \subset N(XA)\ .$

But
$$AXA = A \implies \text{rank}\ AX = \text{rank}\ XA = \text{rank}\, A\ .$$

Theorem. Let X be a $\{1,2\}$–inverses of A. Then:

(a) AX is the projector on $R(A)$ along $N(X)$, and

(b) XA is the projector on $R(X)$ along $N(A)$.

Proof.
$$AX = (AX)^2 \implies AX = P_{R(AX),N(AX)}$$
$$AXA = A \implies R(AX) = R(A)$$
$$XAX = X\ ,\ \text{rank}\, AX = \text{rank}\, X \implies N(AX) = N(X)$$

The set of $\{1,3\}$–inverses

Theorem. The set $A\{1,3\}$ consists of all solutions for X of

$$AX = AA^{(1,3)} \; , \tag{1}$$

where $A^{(1,3)}$ is an arbitrary element of $A\{1,3\}$.

Proof. If X satisfies (1), then

$$AXA = AA^{(1,3)}A = A \; , \; AX = (AX)^* \; . \quad \therefore \; X \in A\{1,3\} \; .$$

Conversely, if $X \in A\{1,3\}$, then

$$\begin{aligned} AA^{(1,3)} &= AXAA^{(1,3)} = (AX)^*AA^{(1,3)} = X^*A^*(A^{(1,3)})^*A^* \\ &= X^*A^* = AX \; . \end{aligned}$$

Theorem. The set $A\{1,4\}$ consists of all solutions for X of

$$XA = A^{(1,4)}A \; .$$

Characterizations of $\{1,3\}$, and $\{1,4\}$–inverses

Recall that for $\mathbb{C}^n = L \oplus M$.

$$M = L^{\perp} \iff P_{L,M} \text{ is Hermitian}$$

Theorem. For any $A \in \mathbb{C}^{m\times n}$:

(a) $AX = P_{R(A)} \iff X \in A\{1,3\}$

(b) $XA = P_{R(A^*)} \iff X \in A\{1,4\}$

Proof. (a) $\Longleftarrow$

$$AXA = A \implies AX = P_{R(AX),N(AX)}$$

$$AXA = A \implies R(AX) = R(A) \ \therefore\ AX = P_{R(A),N(AX)}$$

$$AX = (AX)^* \implies N(AX) = R(A)^{\perp} \ \therefore\ AX = P_{R(A)}$$

(a) $\Longrightarrow$

$$AX = P_{R(A)} = AA^{(1,3)} \implies X \in A\{1,3\}$$

$\{1,2,3\}$, and $\{1,2,4\}$–inverses

Theorem (Urquhart). For every $A \in \mathbb{C}^{m\times n}$,

$$(A^*A)^{(1)}A^* \in A\{1,2,3\}\ , \tag{a}$$

$$A^*(AA^*)^{(1)} \in A\{1,2,4\}\ , \tag{b}$$

$$A^{(1,4)}AA^{(1,3)} \in A\{1,2,3,4\}\ . \tag{c}$$

Proof of (a). Let $X := (A^*A)^{(1)}A^*$.

$$R(A^*A) = R(A^*) \text{ (why?)} \implies A^* = A^*AU\ ,\ \exists\, U \quad \therefore\ A = U^*A^*A$$

$$\therefore\ AXA = U^*A^*A(A^*A)^{(1)}A^* = U^*A^*A = A \quad \therefore\ X \in A\{1\}$$

$$\operatorname{rank} X \le \operatorname{rank} A^* \text{ and } X \in A\{1\} \implies \operatorname{rank} X \ge \operatorname{rank} A$$

$$\therefore\ \operatorname{rank} X = \operatorname{rank} A \quad \therefore\ X \in A\{2\}$$

Finally

$$AX = U^*A^*A(A^*A)^{(1)}A^*AU = U^*A^*AU \quad \therefore X \in A\{3\}\ \square$$

The Moore–Penrose inverse

Theorem (Penrose). Given $A \in \mathbb{C}^{m\times n}$, a solution of

$$AXA = A \,, \tag{1}$$

$$XAX = X \,, \tag{2}$$

$$(AX)^* = AX \,, \tag{3}$$

$$(XA)^* = XA \,, \tag{4}$$

exists and is unique. The $\{1,2,3,4\}$–inverse of A is denoted $A^\dagger$.

Proof. Uniqueness. Let $X, Y \in A\{1,2,3,4\}$. Then

$$\begin{aligned} X &= X(AX)^* = XX^*A^* = X(AX)^*(AY)^* \\ &= XAY = (XA)^*(YA)^*Y = A^*Y^*Y \\ &= (YA)^*Y = Y \,. \end{aligned}$$

Existence. $A^\dagger = A^{(1,4)}AA^{(1,3)}$. □

Full–rank factorization

Given $A \in \mathbb{C}_r^{m\times n}$, $r > 0$, a **full–rank factorization** is

$$A = CR \,,\; C \in \mathbb{C}_r^{m\times r} \,,\; R \in \mathbb{C}_r^{r\times n} \tag{1}$$

Theorem (MacDuffee). Given $A \in \mathbb{C}_r^{m\times n}$, $r > 0$, C, R as in (1),

$$A^\dagger = R^*(C^*AR^*)^{-1}C^* \,. \tag{2}$$

Proof. C^*AR^* is nonsingular, because

$$C^*AR^* = (C^*C)(RR^*) \,,\; \text{a product of nonsingular matrices} \,.$$

Let $X =$ RHS(2) $= R^*(RR^*)^{-1}(C^*C)^{-1}C^*$, and check that X satisfies the 4 Penrose equations. □

$$A^\dagger = R^*(RR^*)^{-1}(C^*C)^{-1}C^* = R^\dagger C^\dagger \tag{3}$$

Q: What is a "good" method for full–rank factorization ?

Singular value decomposition

Let $A \in \mathbb{C}_r^{m\times n}$, $r > 0$, and let

$$AA^*\mathbf{u}_i = \sigma_i^2\,\mathbf{u}_i\ ,\ \ i \in \overline{1,m}$$

$$A^*A\mathbf{v}_i = \sigma_i^2\,\mathbf{v}_i\ ,\ \ i \in \overline{1,n}$$

$$\sigma_1 \geq \sigma_1 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \sigma_{r+2} = \cdots$$

The **singular value decomposition** (**SVD**) of A is

$$A = U\Sigma V^* \tag{SVD}$$

$$U = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_m] \in \mathbb{C}^{m\times m}\ ,\ U^*U = I_m\ ,$$

$$V = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n] \in \mathbb{C}^{n\times n}\ ,\ V^*V = I_n\ ,$$

$$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r) \in \mathbb{R}^{m\times n}\ .$$

Theorem (Penrose).

$$A^\dagger = V\Sigma^\dagger U^*$$

$$\text{where } \Sigma^\dagger = \mathrm{diag}\left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \cdots, \frac{1}{\sigma_r}\right) \in \mathbb{R}^{n\times m}$$

Properties of the Moore–Penrose inverse

(a) For any scalar λ, $$\lambda^\dagger = \begin{cases} \frac{1}{\lambda}\,, & \text{if } \lambda \neq 0\,; \\ 0\,, & \text{otherwise}\,. \end{cases}$$

If $\mathbf{a}, \mathbf{b}$ are column vectors then

(b) $\mathbf{a}^\dagger = (\mathbf{a}^*\mathbf{a})^\dagger \mathbf{a}^*$

(c) $(\mathbf{a}\mathbf{b}^*)^\dagger = (\mathbf{a}^*\mathbf{a})^\dagger (\mathbf{b}^*\mathbf{b})^\dagger \mathbf{b}\mathbf{a}^*$

(d) If $D = \text{diag}(\lambda_1, \cdots, \lambda_k) \in \mathbb{C}^{m\times n}$ then
$$D^\dagger = \text{diag}(\lambda_1^\dagger, \cdots, \lambda_k^\dagger) \in \mathbb{C}^{n\times m}$$

For any matrix A

(e) $(A^\dagger)^\dagger = A$

(f) $(A^*)^\dagger = (A^\dagger)^*$

(g) $(A^T)^\dagger = (A^\dagger)^T$

(h) $A^\dagger = (A^*A)^\dagger A^* = A^*(AA^*)^\dagger$

(i) $R(A^\dagger) = R(A^*)$

(j) $N(A^\dagger) = N(A^*)$

(k) $AA^\dagger = P_{R(A)}$

(l) $A^\dagger A = P_{R(A^*)}$

(m) If U and V are unitary matrices, $(UAV)^\dagger = V^*A^\dagger U^*$

(n) For any matrices A, B: $(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$

Non–properties of the Moore–Penrose inverse

(a) In general, for compatible A, B,

$$(AB)^\dagger \neq B^\dagger A^\dagger$$

(b) If A, B are similar, i.e. $B = S^{-1}AS$ for some nonsingular S, then, in general, $B^\dagger \neq S^{-1}A^\dagger S$.

(c) If $J_k(0)$ is a Jordan block corresponding to the eigenvalue zero, then $(J_k(0))^\dagger = (J_k(0))^T$. For example,

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}^\dagger = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$\therefore$ $A^\dagger$ is not a polynomial in A.

Continuity of the inverse

Let $\|\cdot\|$ be a **multiplicative matrix norm**, i.e.

$$\|XY\| \leq \|X\|\|Y\| \text{ , if } XY \text{ is defined}$$

Let $X \in \mathbb{C}_n^{n\times n}$. Then the **perturbation** $(X+E) = (I + EX^{-1})X$ is **nonsingular** for all E such that $\|E\| < \dfrac{1}{\|X^{-1}\|}$ and its inverse is

$$(X+E)^{-1} = X^{-1}\left(I - EX^{-1} + (EX^{-1})^2 - (EX^{-1})^3 + \cdots\right)$$

which converges if

$$\|EX^{-1}\| < 1 \text{ , guaranteed by } \|E\| < \frac{1}{\|X^{-1}\|}$$

The **inverse** is a **continuous function** $\mathbb{C}_n^{n\times n} \mapsto \mathbb{C}_n^{n\times n}$, and the **nonsingular matrices** are an **open set** in $\mathbb{C}^{n\times n}$.

The Moore–Penrose inverse is discontinuous

Ex. Let

$$X(\epsilon) = \begin{bmatrix} 1 & 0 \\ 0 & \epsilon \end{bmatrix} \to X(0) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \text{ as } \epsilon \to 0 \,.$$

But

$$X(\epsilon)^\dagger = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\epsilon} \end{bmatrix} \not\to X(0)^\dagger = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} .$$

For perturbations $E_k \to O$,

$$(X + E_k)^\dagger \to X^\dagger \iff \text{rank}\,(X + E_k) \to \text{rank}\, X$$

The Smith normal form

A nonsingular matrix $A \in \mathbb{Z}^{n \times n}$ whose inverse A^{-1} is also in $\mathbb{Z}^{n \times n}$ is called a **unit matrix**.

Two matrices $A, S \in \mathbb{Z}^{m \times n}$ are said to be **equivalent over** $\mathbb{Z}$ if there exist two unit matrices $P \in \mathbb{Z}^{m \times m}$ and $Q \in \mathbb{Z}^{n \times n}$ such that

$$PAQ = S \, . \tag{1}$$

Theorem. Let $A \in \mathbb{Z}_r^{m \times n}$. Then A is equivalent over $\mathbb{Z}$ to a matrix $S = [s_{ij}] \in \mathbb{Z}_r^{m \times n}$ such that:

(a) $s_{ii} \neq 0 \,, \quad i \in \overline{1, r}$,

(b) $s_{ij} = 0$ otherwise, and

(c) s_{ii} divides $s_{i+1,i+1}$ for $i \in \overline{1, r-1}$.

S is called the **Smith normal form** of A, and its nonzero elements s_{ii} $(i \in \overline{1, r})$ are **invariant factors** of A.

Integer solutions

Let $A \in \mathbb{Z}^{m\times n}, \mathbf{b} \in \mathbb{Z}^m$ and let the linear equation

$$A\mathbf{x} = \mathbf{b} \tag{P}$$

be consistent. It is required to determine if (P) has an integer solution, in which case determine all of them.

Theorem (**Hurt and Waid**). Let $A \in \mathbb{Z}^{m\times n}$. Then there is an $n \times m$ matrix X satisfying

$$AXA = A\ , \tag{1}$$

$$XAX = X\ , \tag{2}$$

$$AX \in \mathbb{Z}^{m\times m}, \quad XA \in \mathbb{Z}^{n\times n}\ . \tag{6}$$

Proof. Let $PAQ = S$ be the Smith normal form of A. Then

$$X = QS^{\dagger}P\ .$$

Integer solutions (con'd)

Let $\widehat{A}$ the $\{1,2\}$–inverse of A as given above.

Theorem (**Hurt and Waid**). Let A and $\mathbf{b}$ be integral, and let the vector equation

$$A\mathbf{x} = \mathbf{b} \tag{P}$$

be consistent. Then (P) has an integral solution if and only if the vector

$$\widehat{A}\mathbf{b}$$

is integral, in which case the general integral solution of (P) is

$$\mathbf{x} = \widehat{A}\mathbf{b} + (I - \widehat{A}A)\mathbf{y} \ , \quad \mathbf{y} \in \mathbb{Z}^n \ .$$

Application of {2}–inverses to Newton's method

The **Newton method** for solving a single equation in 1 variable,

$$f(x) = 0 \ ,$$

is

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)} \ , \ (k = 0, 1, \ldots) \ .$$

A **Newton method** for solving m equations in n variables

$$f_i(x_1, \ldots, x_n) = 0 \ , \ i \in \overline{1, m} \quad \text{or} \quad \mathbf{f}(\mathbf{x}) = \mathbf{0} \ ,$$

is similarly given, for the case $m = n$, by

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \mathbf{f}'(\mathbf{x}^k)^{-1}\mathbf{f}(\mathbf{x}^k) \ , \ (k = 0, 1, \ldots) \ ,$$

where $\mathbf{f}'(\mathbf{x}^k)$ is the **derivative** of $\mathbf{f}$ at $\mathbf{x}^k$, represented by the matrix of partial derivatives (the **Jacobian** matrix)

$$\mathbf{f}'(\mathbf{x}^k) = \left(\frac{\partial f_i}{\partial x_j}(\mathbf{x}^k) \right) \ .$$

Notation

We denote the **derivative** of $\mathbf{f}$ at $\mathbf{c}$

$$\mathbf{f}'(\mathbf{c}) = \left(\frac{\partial f_i}{\partial x_j}(\mathbf{c})\right) \text{ by } J_{\mathbf{f}}(\mathbf{c}) \text{ or by } J_{\mathbf{c}}\ .$$

We denote by $\|\cdot\|$ both a vector norm in $\mathbb{R}^n$ and a matrix norm consistent with it,

$$\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|\ ,\ \forall\ \mathbf{x}\ .$$

For a given point $\mathbf{x}^0 \in \mathbb{R}^n$ and a positive scalar r we denote by

$$B(\mathbf{x}^0, r) = \{\mathbf{x} \in \mathbb{R}^n : \| \mathbf{x} - \mathbf{x}^0 \| < r\}$$

the **open ball** with **center** $\mathbf{x}^0$ and **radius** r. The **closed ball** with the same center and radius is

$$\overline{B(\mathbf{x}^0, r)} = \{\mathbf{x} \in \mathbb{R}^n : \| \mathbf{x} - \mathbf{x}^0 \| \leq r\}\ .$$

Newton method using $\{2\}$–inverses of $\mathbf{f}'$

Theorem. Let $\mathbf{x}^0 \in \mathbb{R}^n$, $r > 0$ and let $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ be differentiable in $B(\mathbf{x}^0, r)$. Let $M > 0$ be such that

$$\|J_{\mathbf{u}} - J_{\mathbf{v}}\| \leq M \|\mathbf{u} - \mathbf{v}\| \tag{1}$$

for all $\mathbf{u}, \mathbf{v} \in B(\mathbf{x}^0, r)$. Further, assume that for all $\mathbf{x} \in \overline{B(\mathbf{x}^0, r)}$, the Jacobian $J_{\mathbf{x}}$ has a $\{2\}$–inverse $T_{\mathbf{x}} \in \mathbb{R}^{n \times m}$, $T_{\mathbf{x}} J_{\mathbf{x}} T_{\mathbf{x}} = T_{\mathbf{x}}$,

$$\text{such that} \quad \|T_{\mathbf{x}^0}\| \left\|\mathbf{f}(\mathbf{x}^0)\right\| < \alpha \,, \tag{2}$$

$$\text{and,} \qquad \|(T_{\mathbf{u}} - T_{\mathbf{v}})\mathbf{f}(\mathbf{v})\| \leq N \|\mathbf{u} - \mathbf{v}\|^2 \,, \ \forall\, \mathbf{u}, \mathbf{v} \in B(\mathbf{x}^0, r) \tag{3}$$

$$\frac{M}{2} \|T_{\mathbf{u}}\| + N \leq K < 1 \,, \ \forall\, \mathbf{u} \in B(\mathbf{x}^0, r) \tag{4}$$

for some positive scalars N, K and α, and

$$h := \alpha K < 1 \,, \ \frac{\alpha}{1 - h} < r \,. \tag{5}$$

Theorem (cont'd)

Then:

(a) Starting at $\mathbf{x}^0$, all iterates

$$\mathbf{x}^{k+1} = \mathbf{x}^k - T_{\mathbf{x}^k}\mathbf{f}(\mathbf{x}^k), \;\; k = 0, 1, \ldots \tag{6}$$

lie in $B(\mathbf{x}^0, r)$.

(b) The sequence $\{\mathbf{x}^k\}$ converges, as $k \to \infty$, to a point $\mathbf{x}^\infty \in \overline{B(\mathbf{x}^0, r)}$, that is a solution of

$$T_{\mathbf{x}^\infty}\mathbf{f}(\mathbf{x}) = 0 \; . \tag{7}$$

(c) For all $k \geq 0$

$$\left\|\mathbf{x}^k - \mathbf{x}^\infty\right\| \leq \alpha \frac{h^{2^k - 1}}{1 - h^{2^k}} \; . \tag{8}$$

Since $0 < h < 1$, the method is (at least) quadratically convergent.

The iterates converge not to a solution of $\mathbf{f}(\mathbf{x}) = \mathbf{0}$, but of (7). The degree of approximation depends on the $\{2\}$–inverse used.