THE CONJUGATE GRADIENT METHOD:
AN ITERATIVE METHOD FOR SOLVING $Ax = b$

1. INTRODUCTION

$$\underset{\substack{n \times n \\ \text{REAL MATRIX}}}{A}\; x = \underset{\substack{n \times 1 \text{ REAL COLUMN VECTOR}}}{b} \qquad (1.1)$$

- CAN BE SOLVED DIRECTLY BY LU DECOMPOSITION
  - STRAIGHTFORWARD APPROACH REQUIRES:

    $O(n^2)$ STORAGE

    $O(n^3)$ FLOPS

- FOR LARGE n, AND SPARSE A MORE EFFICIENT APPROACHES EXIST.

  e.g., THE CONJUGATE GRADIENT METHOD [HESTENES & STIEFEL, 1952]

SOME NOTATION:

A, H	REAL, SYMMETRIC, POSITIVE-DEFINITE $n \times n$ MATRICES
x, r, p, b, y	REAL $n \times 1$ COLUMN VECTORS
A^T	SIGNIFIES THE TRANSPOSE OF THE MATRIX A
$x^{(k)}$	SIGNIFIES THE k-TH ITERATE OF x
x_k	SIGNIFIES THE k-TH COMPONENT OF VECTOR x
α, β	REAL SCALARS

2. THE BASIC CG METHOD

a. MINIMIZING A FUNCTION

LET

$$F = \frac{1}{2} x^T A x - x^T b \qquad (2.1)$$

AND $x_0 = A^{-1} b$

THEN, WE CAN RE-WRITE (2.1) AS:

$$F = \underbrace{\frac{1}{2} (x-x_0)^T A (x-x_0)}_{\substack{\text{MINIMIZED BY } x=x_0 \\ \text{SINCE } A \text{ POSITIVE-} \\ \text{DEFINITE} \\ (\text{i.e. } y^T A y > 0 \ \forall\, y \neq 0)}} - \underbrace{\frac{1}{2} x_0^T A x_0}_{\substack{\text{CONSTANT W.R.T.} \\ x}}$$

$\Rightarrow$ F MINIMIZED AT $x = x_0 = A^{-1} b$

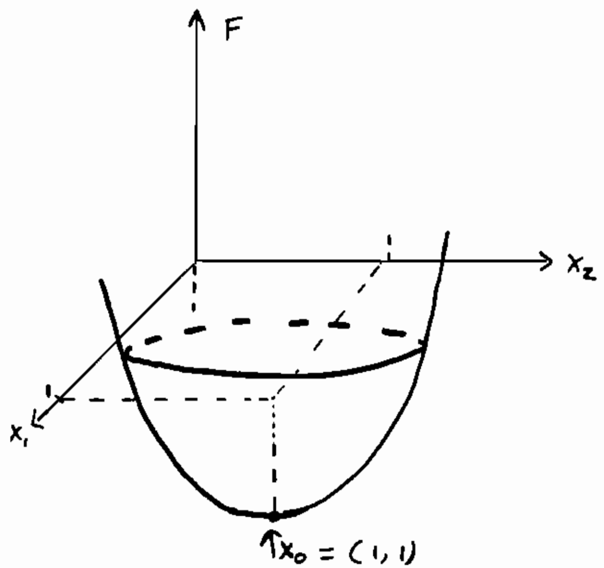
i.e., FINDING THE <u>MINIMUM POINT OF F</u> WILL GIVE THE <u>SOLUTION TO EQN. (1.1)</u>

EXAMPLE ($n=2$, i.e., 2-DIMENSIONAL)

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \; ; \quad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \Rightarrow \quad x_0 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

ALSO, $F = \frac{1}{2}(x_1 - 1)^2 + (x_2 - 1)^2 - \frac{3}{2}$

FIGURE 1



$\Rightarrow$ THE MINIMUM OF F AND THE SOLUTION TO $Ax = b$ IS THE VECTOR $x_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

b. SEARCHING

CG METHOD USES A LINEAR SEARCH TO FIND THE MINIMUM OF F AS FOLLOWS:

- GIVEN AN INITIAL GUESS $x^{(0)}$ AND A SEQUENCE OF VECTORS $p^{(0)}, p^{(1)}, \ldots$ CALLED SEARCH DIRECTIONS, FIND THE <u>MINIMUM</u> OF F ALONG EACH SEARCH DIRECTION IN TURN.

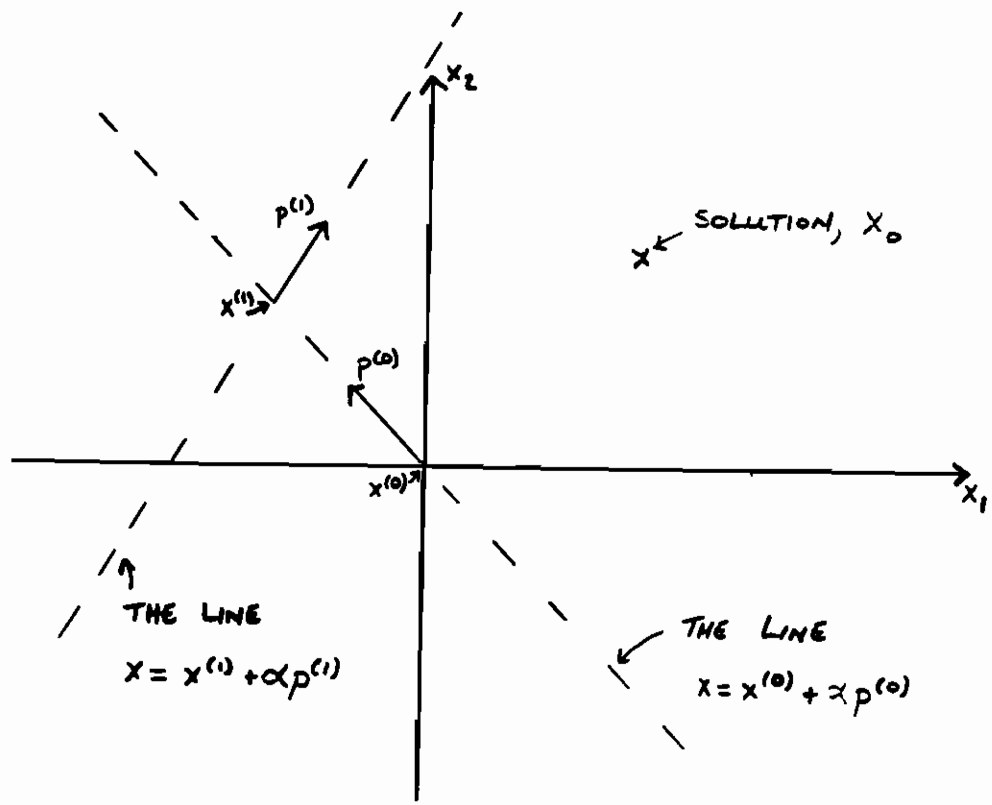
i.e., FIND A NEW ESTIMATE $x^{(k+1)}$ FROM AN OLD ESTIMATE $x^{(k)}$ BY SEARCHING FOR THE MINIMUM OF $F(x)$ ALONG THE LINE:

$$x = x^{(k)} + \alpha p^{(k)}$$

SCALAR PARAMETER $(-\infty < \alpha < +\infty)$

HOW ARE THE SEARCH DIRECTIONS CHOSEN? (LATER!)

FIGURE 2



HOW DO WE FIND THE <u>MINIMUM POINT</u> $x^{(k+1)}$ ALONG THE LINE

$$x = x^{(k)} + \alpha p^{(k)} \quad ?$$

INTRODUCE

$$\nabla F = \begin{pmatrix} \partial F/\partial x_1 \\ \partial F/\partial x_2 \\ \vdots \\ \partial F/\partial x_n \end{pmatrix}$$

NOW, FROM EQN. (2.1) :

$$F = \frac{1}{2} \sum_i \sum_j x_i A_{ij} x_j - \sum_i x_i b_i$$

DIFFERENTIATING, WE FIND:

$$\frac{\partial F}{\partial x_k} = \frac{1}{2} \sum_j A_{kj} x_j + \frac{1}{2} \sum_i x_i A_{ik} - b_k$$

SYMMETRY

$$= \frac{1}{2} \sum_j A_{kj} x_j + \frac{1}{2} \sum_j A_{kj} x_j - b_k$$

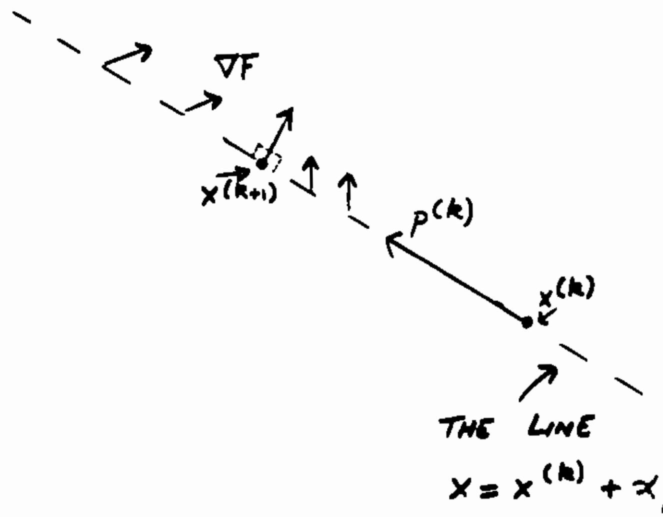
$$= \sum_j A_{kj} x_j - b_k$$

OR, IN MATRIX FORM :

$$\nabla F = Ax - b$$

NOW, AT THE <u>MINIMUM POINT</u>, THE RATE OF CHANGE OF F ALONG $p^{(k)}$ IS <u>ZERO</u>, i.e.,

$$(\nabla F)^T p^{(k)} = 0$$



LET THE VECTOR

$$\boxed{r = b - Ax}$$

BE CALLED THE RESIDUAL, THEN :

$$\nabla F = -r \qquad (2.2)$$

HENCE, AT THE <u>MINIMUM POINT</u> $x^{(k+1)}$, THE RESIDUAL IS $r^{(k+1)} = b - Ax^{(k+1)}$, SO :

$$(r^{(k+1)})^T p^{(k)} = 0 \qquad (2.3)$$

BUT,

$$\boxed{x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)}} \qquad (2.4)$$

LEFT-MULTIPLY BOTH SIDES OF EQN. (2.4) BY A AND SUBTRACT RESULT FROM b :

$$\underbrace{b - Ax^{(k+1)}}_{r^{(k+1)}} = \underbrace{b - (Ax^{(k)}}_{r^{(k)}} + \alpha^{(k)} A p^{(k)})$$

$$\Rightarrow \quad r^{(k+1)} = r^{(k)} - \alpha^{(k)} A p^{(k)} \qquad (2.5)$$

NOW, USING EQN. (2.3) :

$$p^{(k)T} r^{(k+1)} = 0 = p^{(k)T} r^{(k)} - \alpha^{(k)} p^{(k)T} A p^{(k)}$$

RE-ARRANGING :

$$\boxed{\alpha^{(k)} = \frac{p^{(k)T} r^{(k)}}{p^{(k)T} A p^{(k)}}} \qquad (2.6)$$

* EQNS. (2.4) AND (2.6) TELL US PRECISELY WHERE THE <u>NEW MINIMUM POINT</u> $x^{(k+1)}$ LIES.

(<u>N.B.</u> $r^{(k)} = b - Ax^{(k)}$)

EXAMPLE: LET $x^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ AND TAKE TWO ARBITRARY SEARCH DIRECTIONS, SAY $p^{(0)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ AND $p^{(1)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. THEN WE OBTAIN THE FOLLOWING RESULTS:

	$x^{(k)T}$	$p^{(k)T}$	$r^{(k)T}$	$\alpha^{(k)}$
$k=0$	$(0,0)$	$(-1,1)$	$(1,2)$	$+1/3$
$k=1$	$(-1/3, 1/3)$	$(1,2)$	$(4/3, 4/3)$	$+4/9$
$k=2$	$(1/9, 11/9)$	-	-	-
$\vdots$				

(SEE FIG. 2 FOR GRAPHICAL (NOT TO SCALE) RESULTS)

N.B. IF WE CHOOSE SEARCH DIRECTIONS ARBITRARILY LIKE THIS, THE ALGORITHM MAY NEVER CONVERGE!

C. THE METHOD OF STEEPEST DESCENT

AN "INTUITIVE" WAY TO CHOOSE SEARCH DIRECTIONS IS TO ALWAYS GO "DOWNHILL" ON THE STEEPEST PATH:

$$p^{(k)} = -\nabla F(x^{(k)}) = r^{(k)} \qquad (2.7)$$

$\therefore$, FROM EQN. (2.3):

$$r^{(k)T} p^{(k-1)} = 0$$

$$\Rightarrow p^{(k)T} p^{(k-1)} = 0$$

i.e., $p^{(k)}$ ALWAYS $\perp$ TO $p^{(k-1)}$

N.B. $\alpha^{(k)} = \dfrac{r^{(k)T} r^{(k)}}{r^{(k)T} A\, r^{(k)}}$

AND, $x^{(k+1)} = x^{(k)} + \alpha^{(k)} r^{(k)}$

<u>EXAMPLE</u> LET $x^{(0)T} = (0,0)$ AND APPLY THE STEEPEST DESCENT ALGORITHM :

	$x^{(k)T}$	$r^{(k)T}$	$p^{(k)T}$	$\alpha^{(k)}$	$\lvert X_0 - x^{(k)} \rvert$
$k=0$	$(0,0)$	$(1,2)$	$(1,2)$	$5/9$	1.4142
$k=1$	$(5/9, 10/9)$	$(4/9, -2/9)$	$(4/9, -2/9)$	$5/6$	0.4581
$k=2$	$(25/27, 25/27)$	$(2/27, 2/27)$	$(2/27, 2/27)$	$5/9$	0.1048
$k=3$	$(0.963, 1.0082)$				0.0339

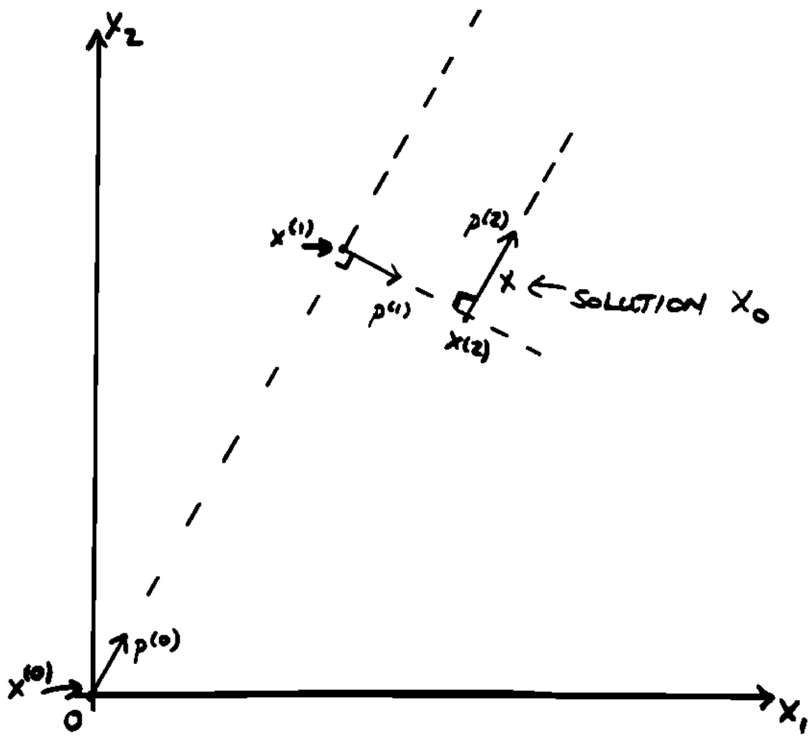
(SEE FIG. 3)

THIS METHOD BRINGS US <u>NEARER & NEARER</u> TO THE SOLUTION X_0 WITH <u>EACH ITERATION</u>.

HOWEVER, 3 LINEAR SEARCHES HAVE FAILED TO PRODUCE THE EXACT ANSWER (EVEN FOR $n=2$, i.e., 2-DIMENSION PROBLEM).

IDEALLY, WOULD PREFER AN ALGORITHM THAT IS <u>GAURANTEED</u> TO FIND THE <u>SOLUTION IN n-STEPS</u>. (BUT NOT IF IT TAKES $O(n)$ TO GET ANYWHERE NEAR THE ANSWER.)

<u>FIGURE 3</u> : STEEPEST DESCENT METHOD



d. CONJUGATE DIRECTION METHODS

INTRODUCE THE CONCEPT OF CONJUGACY:

TWO VECTORS a AND b ARE CONJUGATE (W.R.T. THE SYMMETRIC MATRIX A) IF:

$$a^T A b = 0$$

i.e., CONJUGACY IS A SORT OF GENERALIZED ORTHOGONALITY.

SUPPOSE WE CHOOSE OUR SEARCH DIRECTIONS $p^{(0)}, p^{(1)}, \ldots$ TO BE MUTUALLY CONJUGATE:

$$p^{(j)T} A p^{(k)} = 0 \quad \text{IF } j \neq k \qquad (2.8)$$

AN ALGORITHM THAT USES SUCH SEARCH DIRECTIONS IS CALLED A CONJUGATE DIRECTION (CD) ALGORITHM.

(CG IS A SPECIAL CASE OF CD.)

WE ALREADY HAVE EQN. (2.3):

$$r^{(k+1)T} p^{(k)} = 0$$

AND IF THE p VECTORS ARE MUTUALLY CONJUGATE, IT TURNS OUT THAT:

$$r^{(k+1)T} p^{(j)} = 0 \quad \forall \; j \leq k \qquad (2.9)$$

* NOT ONLY IS THE NEW RESIDUAL ORTHOGONAL TO THE PREVIOUS SEARCH DIRECTION $p^{(k)}$, IT IS ORTHOGONAL TO ALL PREVIOUS SEARCH DIRECTIONS.

$\Rightarrow$ $x^{(k+1)}$ IS A MINIMUM OF F OVER ALL THE VECTORS GIVEN BY

$$x = x^{(0)} + c_0 p^{(0)} + c_1 p^{(1)} + \ldots + c_k p^{(k)}$$

FOR ALL REAL NUMBERS $c_0, c_1, \ldots, c_k$.

$\Rightarrow$ CONVERGENCE TO CORRECT SOLUTION IS GUARANTEED IN AT MOST n-STEPS!

EXAMPLE: $x^{(0)T} = (0,0)$. LET TWO CONJUGATE DIRECTIONS BE $p^{(0)T} = (1,-1)$ AND $p^{(1)T} = (2,1)$.

(CHECK: $p^{(0)T} A\, p^{(1)} = 0$ ✓)

	$x^{(k)T}$	$r^{(k)T}$	$p^{(k)T}$	$\alpha^{(k)}$	$\lvert x_0 - x^{(k)}\rvert$
$k=0$	$(0,0)$	$(1,2)$	$(1,-1)$	$-1/3$	1.4142
$k=1$	$(-1/3, 1/3)$	$(4/3, 4/3)$	$(2,1)$	$2/3$	1.4907
$k=2$	$(1,1)$	$(0,0)$	–	–	0.0000

(SEE FIG. 4)

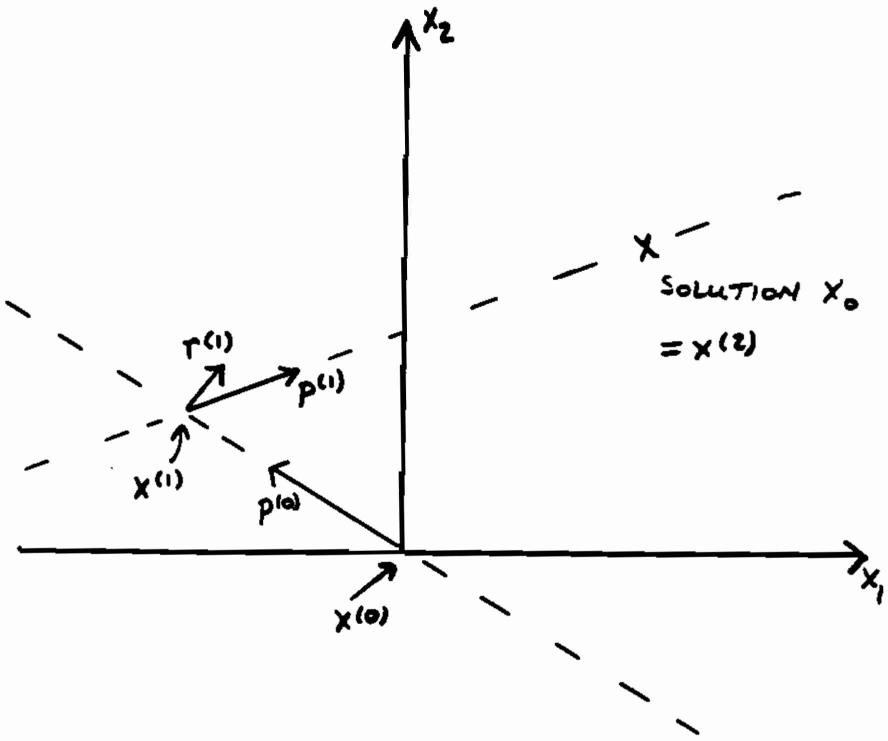
(CHECK: $r^{(k+1)T} p^{(j)} = 0 \quad \forall j \le k$; e.g. $r^{(1)T} p^{(0)} = 0$ ✓)

NOTE: BECAUSE $p^{(0)}$ AND $p^{(1)}$ ARE CONJUGATE, THE TWO CONSECUTIVE MINIMIZATIONS ALONG THE TWO LINES SHOWN ARE EQUIVALENT TO A SINGLE MINIMIZATION OVER THE PLANE:

$$x = x^{(0)} + c_0 p^{(0)} + c_1 p^{(1)}$$

A MINIMIZATION THAT NECESSARILY PRODUCES THE FINAL ANSWER x_0 IN THE $n=2$ CASE.

FIGURE 4: CONJUGATE DIRECTION METHOD



$\boxed{p^{(0)T} A\, p^{(1)}}$ CONJUGACY

TWO PROBLEMS WITH CD :

i) ALGORITHM CONVERGED IN 2 STEPS, BUT ESTIMATE $x^{(1)}$ AFTER 1ST STEP VERY POOR.

IF THE p VECTORS ARE CHOSEN ARBITRARILY (EXCEPT THAT THEY ARE CONJUGATE) THERE IS NO REASON SUCCESSIVE ESTIMATES SHOULD GET CLOSER AND CLOSER TO x_0.

⇒ WE MAY HAVE TO WAIT FOR THE n-TH STEP TO OBTAIN AN ACCEPTABLY CLOSE ESTIMATE.

ii) NO METHOD HAS BEEN SUGGESTED FOR PRODUCING n CONJUGATE SEARCH DIRECTIONS.

THE CG METHOD SOLVES BOTH PROBLEMS (i) & (ii).

e. THE CONJUGATE GRADIENT METHOD

COMBINES THE ADVANTAGES OF STEEPEST DESCENT & CONJUGATE DIRECTION METHODS.

⇒ GUARANTEED TO CONVERGE IN n-STEPS, AND SUCCESSIVE ESTIMATES GET CLOSER AND CLOSER TO THE TRUE SOLUTION.

ESSENTIALLY, THE CG METHOD IS A CD METHOD IN WHICH WE FIND THE SEARCH DIRECTIONS AS FOLLOWS :

$$p^{(0)} = r^{(0)} = b - Ax^{(0)}$$

$$\boxed{p^{(k+1)} = \underbrace{r^{(k+1)}}_{\text{"GRADIENT"}} + \underbrace{\beta^{(k)} p^{(k)}}_{\text{"CONJUGATE"}} \qquad (2.10)}$$

(JUST THE TERM USED IN STEEPEST DESCENT) (ADDED SO THAT $p^{(k+1)}$ CAN BE MADE CONJUGATE TO $p^{(k)}$)

DEFINE $\beta^{(k)}$ TO ENSURE :

$$p^{(k)T} A\, p^{(k+1)} = 0$$

LEFT-MULTIPLY EQN. (2.10) BY $p^{(k)T}A$:

$$0 = p^{(k)T}A\, r^{(k+1)} + \beta^{(k)} p^{(k)T} A\, p^{(k)}$$

RE-ARRANGING :

$$\boxed{\beta^{(k)} = -\frac{p^{(k)T} A\, r^{(k+1)}}{p^{(k)T} A\, p^{(k)}}} \qquad (2.11)$$

* WITH THIS WAY OF CHOOSING THE SEARCH DIRECTIONS, IT MAY BE SHOWN THAT NOT ONLY IS $p^{(k+1)}$ CONJUGATE TO $p^{(k)}$, BUT

$$p^{(k+1)T} A\, p^{(j)} = 0 \quad \forall\ j \neq k+1$$

SO THAT CG IS INDEED A CD METHOD.

EQNS. (2.10) & (2.11) TOGETHER WITH THE BASIC LINEAR SEARCH EQNS. (2.4) & (2.6), AND THE DEFINITION OF r, DEFINE THE <u>BASIC CG ALGORITHM</u> :

INITIAL VECTORS :

GUESS $\quad x^{(0)}$

SET $\quad r^{(0)} = b - Ax^{(0)}$

$\quad p^{(0)} = r^{(0)}$

THEN, FOR $k = 0, 1, \ldots$

$$\alpha^{(k)} = \frac{p^{(k)T} r^{(k)}}{p^{(k)T} A\, p^{(k)}}$$

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)} p^{(k)}$$

LINEAR SEARCH TO FIND NEW MINIMUM ALONG LINE.

$$r^{(k+1)} = b - A x^{(k+1)}$$

$$\beta^{(k)} = -\frac{p^{(k)T} A\, r^{(k+1)}}{p^{(k)T} A\, p^{(k)}}$$

$$p^{(k+1)} = r^{(k+1)} + \beta^{(k)} p^{(k)}$$

FIND NEW SEARCH DIRECTION.

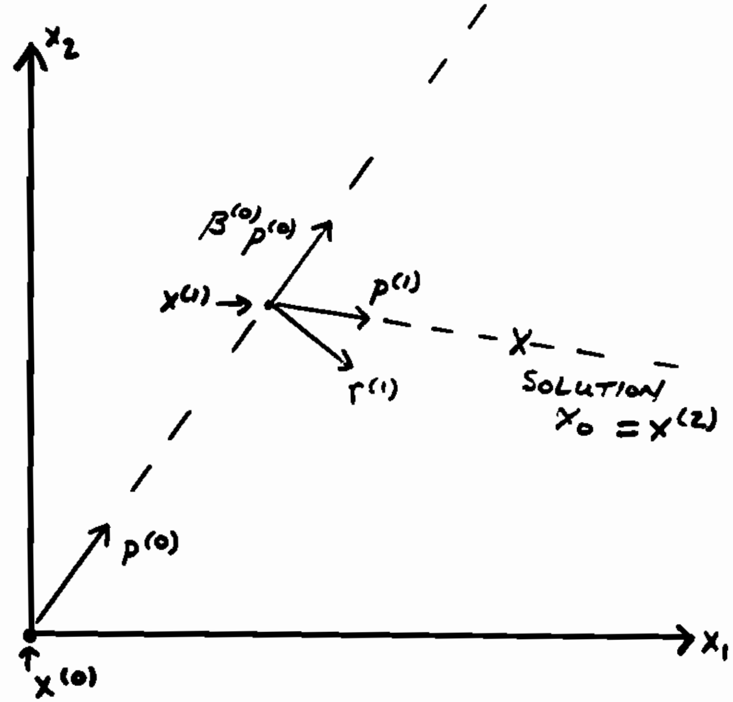
EXAMPLE LET $x^{(0)T} = (0,0)$. APPLYING CG WE FIND:

	$x^{(k)T}$	$r^{(k)T}$	$p^{(k)T}$	$\alpha^{(k)}$	$\lvert x_0 - x^{(k)} \rvert$
$k=0$	$(0,0)$	$(1,2)$	$(1,2)$	$5/9$	1.4142
$k=1$	$(5/9, 10/9)$	$(4/9, -2/9)$	$(0.4938, -0.1235)$	0.9	0.4581
$k=2$	$(1,1)$	$(0,0)$	—	—	0.0000

(SEE FIG. 5)

$\Rightarrow$ CG COMBINES ADVANTAGES OF A CD METHOD AND A STEEPEST DESCENT METHOD.

FIGURE 5: CONJUGATE GRADIENT METHOD



THREE BASIC PROPERTIES OF THE ALGORITHM ARE :

i) CONJUGACY OF THE SEARCH DIRECTIONS

$$p^{(k)T} A \, p^{(j)} = 0 \quad \text{IF} \quad k \neq j$$

ii) ORTHOGONALITY OF THE k-TH RESIDUAL TO THE PREVIOUS SEARCH DIRECTIONS

$$r^{(k)T} p^{(j)} = 0 \quad \text{IF} \quad j < k$$

iii) ORTHOGONALITY OF THE RESIDUALS

$$r^{(j)T} r^{(k)} = 0 \quad \text{IF} \quad k \neq j.$$

(THE FIRST TWO ARE CD PROPERTIES)

f. ADDITIONAL PROPERTIES OF CG

(i) IF A HAS ONLY m DISTINCT EIGENVALUES, THEN THE METHOD WILL CONVERGE AFTER m ITERATIONS.

(ii) IF THE EIGENVALUES OF A FORM c CLUSTERS, THEN A GOOD ESTIMATE OF THE SOLUTION WILL BE OBTAINED AFTER c ITERATIONS.

⇒ THE SPEED OF CONVERGENCE OF THE CG METHOD DEPENDS ON THE EIGENVALUES OF MATRIX A.

3. PRECONDITIONED CONJUGATE GRADIENTS (PCG)

a. THE IDEA BEHIND PRECONDITIONING

SOLVE

$$HAx = Hb$$

WITH CONJUGATE GRADIENTS.

H IS A PRECONDITIONING MATRIX SUCH THAT

HA HAS MORE "CLUSTERED" EIGENVALUES THAN A.

$\Rightarrow$ FASTER CONVERGENCE

b. GENERALIZED CG

INSTEAD OF USING

$$p^{(0)} = r^{(0)}$$
$$p^{(k+1)} = r^{(k+1)} + \beta^{(k)} p^{(k)}$$
$$\beta^{(k)} = -\frac{p^{(k)T} A\, r^{(k+1)}}{p^{(k)T} A\, p^{(k)}}$$

IN THE BASIC CG ALGORITHM (SEE p. 22), USE:

$$p^{(0)} = H r^{(0)}$$
$$p^{(k+1)} = H r^{(k+1)} + \beta^{(k)} p^{(k)}$$
$$\beta^{(k)} = -\frac{p^{(k)T} A H\, r^{(k+1)}}{p^{(k)T} A\, p^{(k)}}$$

WHERE H IS A REAL, SYMMETRIC, POSITIVE DEFINITE $n \times n$ MATRIX.

$\Rightarrow$ EIGENVALUE PROPERTIES NOW DEPEND ON HA INSTEAD OF A

EXAMPLE :

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$H = \begin{bmatrix} 1.1 & 0 \\ 0 & 0.5 \end{bmatrix} \qquad HA = \begin{bmatrix} 1.1 & 0 \\ 0 & 1 \end{bmatrix}$$

$\Rightarrow$ H IS CLEARLY A GOOD PRECONDITIONING MATRIX!

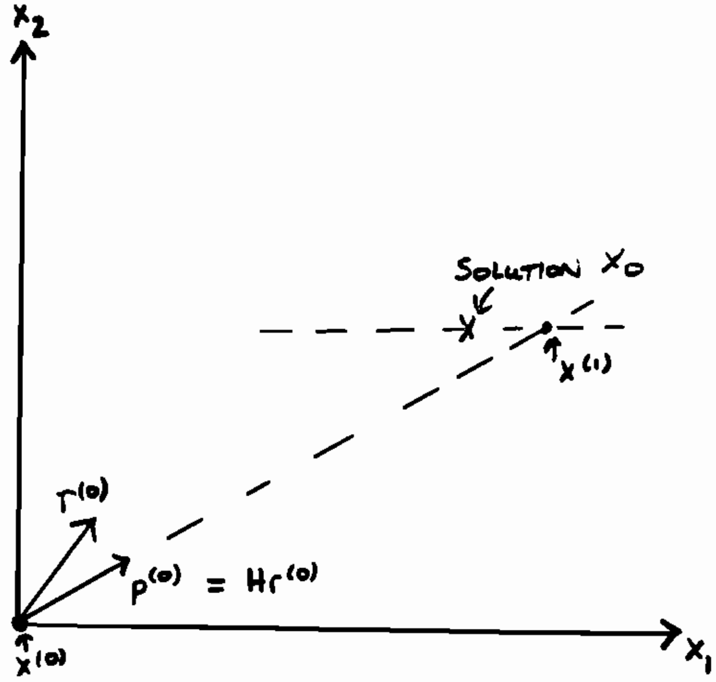
THE GENERALIZED CG METHOD GIVES:

	$x^{(k)T}$	$r^{(k)T}$	$p^{(k)T}$	$\alpha^{(k)}$	$\lvert x_0 - x^{(k)} \rvert$
$k=0$	(0, 0)	(1, 2)	(1.1, 1)	0.9657	1.4142
$k=1$	(1.0623, 0.9657)	(-0.0623, 0.0685)	(-0.0662, 0.0364)	0.9413	0.0711
$k=2$	(1, 1)	(0, 0)	—	—	0.0

(SEE FIGURE 6)

WE COULD STOP AFTER THE FIRST STEP ($k=1$) AND HAVE A GOOD ESTIMATE FOR X.

FIGURE 6



Note: The basic CG algorithm is just the special case $H = I$ (identity)

Suppose, we put $H = A^{-1}$

$$\Rightarrow \quad HA = I$$

which has just 1 eigenvalue, and the algorithm would converge in exactly 1 step. (But computing A^{-1} defeats our purpose!)

$\Rightarrow$ **We need to find an H that:**

(i) is better than just $H = I$, but which

(ii) is less expensive to compute than $H = A^{-1}$, and

(iii) we would like H to be stored in as few locations as possible (recall, A is sparse)

$\Rightarrow$ Use an Incomplete Choleski (IC) Decomposition of A.

c. Incomplete Choleski Decomposition

Recall, for a sparse matrix, the Choleski decomposition produces fill-in:

$$A = \begin{bmatrix} x & x & x & x \\ x & x & & \\ x & & x & \\ x & & & x \end{bmatrix}$$

$$\Downarrow$$

$$L = \begin{bmatrix} x & & & \\ x & x & & \\ x & x & x & \\ x & x & x & x \end{bmatrix}$$

$\uparrow$ These 3 elements have become non-zero: fill-in

i.e., sparse $A \not\Rightarrow$ sparse L

INCOMPLETE CHOLESKI (IC) DECOMPOSITION OF A SPARSE MATRIX

$\Rightarrow$ USE CHOLESKI ALGORITHM, BUT

$$L_{ij} = 0 \quad \text{IF} \quad A_{ij} = 0$$

i.e., IMPOSE THE SAME SPARSITY PATTERN ON L AS NATURALLY EXISTS FOR A.

DECOMPOSITION WILL BE LESS EXPENSIVE & THE RESULTING L MAY BE COMPACTLY STORED).

BUT, WE NO LONGER HAVE

$$A = LL^T$$

INSTEAD IC GIVES:

$$A = LL^T + E$$

WHERE E IS A (SMALL?) ERROR MATRIX.

$\Rightarrow$ THE DECOMPOSITION IS NO LONGER EXACT, IT IS INCOMPLETE (HENCE, THE NAME IC).

NOW, FOR PCG SET

$$H = (LL^T)^{-1}$$

THEN,
$$HA = (LL^T)^{-1}A = I + (LL^T)^{-1}E$$

WHICH WILL HOPEFULLY BE "CLOSER" TO THE IDENTITY MATRIX THAN A ITSELF.

* IN PRACTICE, PCG USING IC DECOMPOSITION IS FOUND TO WORK WELL

$\Rightarrow$ IC-PCG METHOD CONVERGES IN $O(\sqrt{n})$ STEPS IN MOST CASES.

N.B. H MAY BE COMPUTED BY PERFORMING $(L^T)^{-1} L^{-1}$
(THIS IS RELATIVELY CHEAP SINCE L IS LOWER TRIANGULAR & SPARSE).

THE ALGORITHM FOR IC IS:

FOR $j = 1, \ldots, n$:

$$L_{jj} = +\sqrt{A_{jj} - \sum_{k=1}^{j-1} L_{jk}^2}$$

FOR $i = j+1, \ldots, n$:

IF $A_{ij} \neq 0$

$$L_{ij} = \left(A_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{jk}\right) \Big/ L_{jj}$$

ELSE

$L_{ij} = 0$

$O(ns^2)$ ||| ~2 OR 3 CG STEPS

$O(s^2)$

($\equiv$ # OF NON-ZERO'S PER ROW AVERAGE)

N.B. IF A IS POSITIVE-DEFINITE & DECOMPOSITION IS COMPLETE, THEN

$$A_{jj} > \sum_{k=1}^{j-1} L_{jk}^2$$

BUT, FOR INCOMPLETE DECOMPOSITION, THERE IS NO GAURANTEE THAT THIS CONDITION HOLDS!

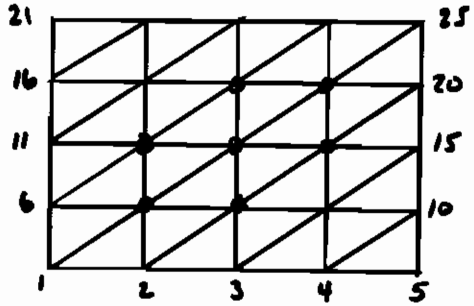
$\Rightarrow$ IF IT DOESN'T, SET L_{jj} TO SMALL +VE NUMBER.

d. PERFORMANCE OF IC-PCG vs. BANDED CHOLESKI

WITH IC-PCG, THERE IS NO NEED TO RE-ORDER THE NODES TO OBTAIN A SMALL BANDWIDTH.

ALSO, ALL PREDICTABLE ZEROS ARE IGNORED BY IC-PCG. (ZEROS IN THE BAND TREATED AS NON-ZEROS BY BANDED CHOL.)

EXAMPLE REGULAR $N \times N$ FIRST-ORDER FE MESH:



$n = N^2$

$(N = 5)$

SEMI-BANDWIDTH $b = 2N + 1$ (FOR OPTIMAL NODE NUMBERING)

NUMBER OF NON-ZEROS PER ROW ≤ 7

(SAY $s = 7$)

PCG FLOPS:

INCOMPLETE CHOLESKI : $O(ns^2)$

ONE CG STEP : $O(ns)$ (EXPLAIN!)

NUMBER OF CG STEPS : $O(\sqrt{n})$

SO, FOR FIXED s, $(s \ll n)$

TOTAL FLOPS $= O(\sqrt{n}\, n\, s) \simeq O(n^{1.5})$

$\cong O(N^3)$

BANDED CHOLESKI FLOPS

DECOMPOSITION: $O(n b^2)$

SO FOR $b = 2N+1$,

TOTAL FLOPS $\cong O(N^2 N^2)$

$= O(N^4)$

4. CONJUGATE GRADIENTS FOR SYMMETRIC COMPLEX OR INDEFINITE MATRICES

NEED TO SOLVE

$$A x = b$$

$n \times n$ SPARSE SYMMETRIC MATRIX

ELECTROSTATICS & MAGNETOSTATICS $\longrightarrow$ A IS REAL, SYMMETRIC AND POSITIVE DEFINITE

TIME HARMONIC PROBLEMS $\longrightarrow$ A IS COMPLEX AND SYMMETRIC

CG CAN BE USED TO FIND THE <u>STATIONARY POINT</u> OF

$$F = \frac{1}{2} x^T A x - x^T b$$

<u>N.B.</u> x^T : TRANSPOSE OF x
<u>NOT</u> COMPLEX-CONJUGATE-TRANSPOSE !

ALONG EACH SEARCH DIRECTION $p^{(k)}$, THE POINT WE NOW WISH TO FIND IS THE STATIONARY POINT OF F RATHER THAN THE MINIMUM, BUT THE CRITERION IS EXACTLY THE SAME, i.e.,

$$(\nabla F)^T \Big|_{x = x^{(k+1)}} p^{(k)} = 0 .$$

HOWEVER, IF

$$p^{(k)^T} A \, p^{(k)} = 0$$

FOR SOME SEARCH DIRECTION $p^{(k)}$, THE (WHY?) <u>ALGORITHM FAILS</u>. i.e., IF A SEARCH DIRECTION IS SELF-CONJUGATE (W.R.T. A), THEN F VARIES LINEARLY IN THAT DIRECTION & THERE WILL BE <u>NO</u> STATIONARY POINT !

(WON'T OCCUR IF A IS POSITIVE-DEFINITE !)

<u>EXAMPLE</u> : CHOOSE INDEFINITE A :

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \qquad \Rightarrow \quad x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

APPLYING CG : $x^{(0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$; $p^{(0)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$\Rightarrow \; p^{(0)^T} A p^{(0)} = 0$ ($p^{(0)}$ SELF-CONJUGATE !)

$F(x^{(0)} + \alpha p^{(0)}) = -2$ (i.e. INDEPENDENT OF α)

IN PRACTICE, EXACT FLOATING-POINT ZERO FOR $p^T A p$ IS UNLIKELY, BUT WILL RESULT IN A VERY SMALL NUMBER

$\Rightarrow$ LARGE ROUNDING ERRORS

$\Rightarrow$ POOR CONVERGENCE.

THIS METHOD FOR SYMMETRIC INDEFINITE AND COMPLEX MATRICES IS A SPECIAL CASE OF THE COMPLEX BI-CONJUGATE GRADIENT ALGORITHM (REF. [6]).

REFERENCES

The original CG method was devised by Hestenes and Stiefel [1] in 1952. However, it became more popular in 1971 when Reid [2] directed attention to its attractive features as a sparse matrix solver. Preconditioning for a certain restricted class of matrices was developed by Meijerink and van der Vorst in 1977 [3], and this was quickly followed by demonstrations of preconditioned CG for general positive definite and indefinite matrices (Kershaw [4], Manteuffel [5]). In recent years there have been numerous papers on PCG - consult, for example, Jacobs [6] for references. Finally, the more recent book by Hestenes [7] is both mathematically rigorous and comprehensive (except in the area of pre-conditioning, which it hardly discusses). Hestenes develops CG in the larger context of algorithms for minimizing quadratic functions - he deals with both descent and CD methods.

[1] Hestenes, M. R. and Stiefel, E. I "Methods of conjugate gradients for solving linear systems", J. Res. Nat. Bur. Standards, vol. 49, 404-436, 1952

[2] Reid, J. K., "On the method of conjugate gradients for the solution of large sparse systems of linear equations", in "Large sparse sets of linear equations", edited by J. K. Reid, Academic Press, London, 1971

[3] MeijerinK, J. A. and van der Vorst, H. A., "An iterative solution method for linear systems of which the coefficient matrix is symmetric M-matrix", Math. of Comp.. vol. 31(137), 148-162, 1977.

[4] Kershaw, D.S., "The incomplete Cholesky - conjugate gradient method for the iterative solution of systems of linear equations", J. Compo Phys., vol. 26, 43-65, 1978

[5] Manteuffel, T.A., "The shifted incomplete Cholesky factorization", Sandia Labs Report, May, 1978.

[6] Jacobs, D.H., "Generalizations of the conjugate gradient method for solving non-symmetric and complex systems of algebraic equations", CERL Lab Report, March 1980.

[7] Hestenes, M. R., "Conjugate direction methods in optimization", Springer-Verlag, 1980.