PICK D_0 ... $\nabla g(x_0)$ OR ANYTHING ELSE YOU LIKE

FEATURE:

A, B MATRICES DISTANCE:

$$d(A,B) = \max\{\|Ax - Bx\| : \|x\| \le 1\}$$

THEOREM 3.4.5

D IS $m \times n$ SATISFYING SECANT CONDITION $Dd_k = g(x_{k+1}) - g(x_k)$

THEN

$$d(D_{k+1}, D_k) \le d(D, D_k)$$

QUESTIONS:

- DOES THE METHOD WORK? YES
- IS D_k APPROXIMATING $\nabla g(x)$? NO

CLAIM:

$$\lim_{k \to \infty} \frac{\|(D_k - \nabla g(x_k))(x^{(k)} - x^*)\|}{\|x_k - x^*\|} = 0$$

SO D_k BEHAVES LIKE $\nabla g(X_k)$
ON "IMPORTANT" $X_k - x^*$ BUT
NOT NESESSARILY EVERYWHERE
=> ANSWER S
YES LIKE NEWTON'S METHOD

USAGE FOR MIN $f(x)$:
$g(x) = \nabla f(x)$ FOR CRITICAL POINTS
LIKE W NEWTON'S ADJUST ϵ_k POSSIBLE.
$$X_{k+1} = X_k - \epsilon_k D_k^{-1} \nabla f(x_k)$$
TROUBLE -> NOT ALWAYS DESC. DIRECTION
: -C [D_k NOT NESESSARILY POS. DEF]

BFGS & DFP

"MAKE BROYDEN A BIT NICER"

WE WANT MINIMIZING METHOD

$$X_{k+1} = X_k - t_k D_k^{-1} \nabla f(x_k)$$

ST

1) D_k POSITIVE DEFINITE

2) SECANT CONDITION SATISFIED:

$$D_{k+1}(X_{k+1} - X_k) = \nabla f(X_{k+1}) - \nabla f(X_k)$$

3) D_{k+1} CAN BE EASILY COMPUTED FROM D_k

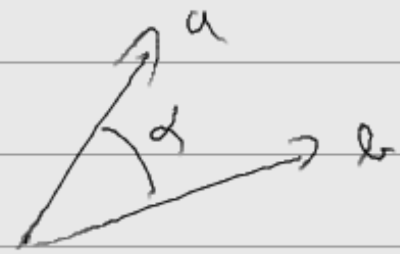
TOOL:

THEOREM 3.5-1

$a, b \in \mathbb{R}^n$ S.T. $a^T b > 0$ THEN EXISTS POSITIVE DEFINITE MATRIX A S.T. $Aa = b$

PROOF:

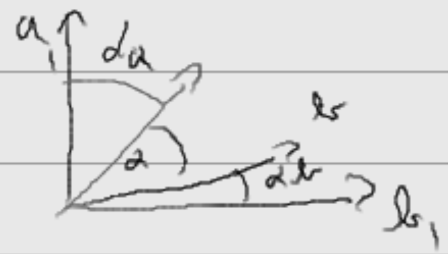
2D SUBSPACE



$\Rightarrow \quad 0 \le \alpha < \pi/2$ SINCE $a^T b > 0$ AND

$$\cos \alpha = \frac{a^T b}{\|a\| \|b\|}$$

CASE I $\quad 0 < \alpha < \pi/2$



PICK a_1 b_1 ORTHOGONAL

S.T. $\alpha_a > 0$

$\alpha_b > 0$

$a = s_1 a_1 + t_1 b_1$

$b = s_2 a_1 + t_2 b_1$

FROM

CONSIDER $A_1 = \begin{pmatrix} \frac{s_2}{s_1} & 0 \\ 0 & \frac{t_2}{t_1} \end{pmatrix}$

WITH BASE a_1 $b_1 = B_1$

$$A_1 (s a_1 + t b_2) = s \left(\frac{s_2}{s_1} \right) a_1 + t \left(\frac{t_2}{t_1} \right) b_1$$

$$\Rightarrow A_1 a_B = b_B$$

CONSIDER BASE $B = \{a, b, c_1, c_2, \ldots, c_{n-2}\}$

$$A = \begin{pmatrix} \frac{\beta_2}{\beta_1} & & & & 0 \\ & \frac{\varepsilon_2}{\varepsilon_1} & & & \\ & & 1 & & \\ & & & 1 & \\ & 0 & & & \ddots\, 1 \end{pmatrix}$$

STILL $Aa = b$ IN BASE

IF ORTHONORMAL U

THEN FOR ORIG.

$$x^T \underbrace{U^T A U}\, x = (Ux) A (Ux) \Rightarrow$$

$\Rightarrow$ $U^T A U$ IS POSITIVE DEFINITE.

CASE II $\alpha = 0 \Rightarrow a = \lambda . b$

- SIMILAR CONSTRUCTION $\rightarrow$ JUST ONE SPECIAL ENTRY INSTEAD OF TWO.

$\square$

$p_k = -D_k^{-1} \nabla f(x_k)$

$d_k = x_{k+1} - x_k$ $\quad$ $x_{k+1} = x_k - t_k \underbrace{D_k^{-1} \nabla f(x_k)}_{p_k}$

$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$

$t_k \in \mathbb{R}$

S.T. CRITERIA $\boxed{1-4}$ SATISFIED OLD

$d_k^T y_k = t_k p_k^T y_k = t_k p_k^T (\nabla f(x_{k+1}) - \nabla f(x_k))$

$= t_k (p_k^T \nabla f(x_{k+1}) - p_k^T \nabla f(x_k))$

3) $\exists\ 0 < \beta < 1$ S.T. $p_k^T \nabla f(x_{k+1}) > \beta p_k^T \nabla f(x_k)$

$> t_k (\beta p_k^T \nabla f(x_k) - p_k^T \nabla f(x_k)) =$

$= t_k (\beta - 1) p_k^T \nabla f(x_k)$

$= \underbrace{t_k}_{>0} \underbrace{(\beta - 1)}_{<0} \Big(\underbrace{-}_{<0} \underbrace{D_k^{-1} \nabla f(x_k))^T \nabla f(x_k)}_{>0} > 0$

AS D_k^{-1} POS DEF

$\Rightarrow \exists\ A_k$ S.T. $A_k d_k = y_k$

RECALL SECANT CONDITION:

$$D_{k+1}(x_{k+1} - x_k) = y(x_{k+1}) - y(x_k)$$

$$D_{k+1}\, d_k = y_k$$

$\rightarrow$ PICK $A_k = D_{k+1}$ WORKS

HOW TO FIND U_k

ST $D_{k+1} = D_k + U_k$?

$U_k = a_k \otimes b_k$ $\leftarrow$ NOT GOOD

IS SYMMETRIC ONLY IF $a_k = \lambda b_k$

AND INDEFINITE IF $\lambda < 0$

$\begin{pmatrix} 1 \\ 2 \end{pmatrix} (2\ 4) = \begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix}$ $\uparrow$ IS POS DEF ?? (NO)

FOR MORE FLEXIBILITY $\downarrow$ $\alpha_k, \beta_k \in \mathbb{R}$

$$U_k = \alpha_k (a_k \otimes a_k) + \beta_k (b_k \otimes b_k)$$

SATISFYING SECANT CONDITION:

$D_{k+1}\, d_k = y_k$

$$y_k = D_k d_k + \alpha_k (a_k \otimes a_k) d_k + \beta_k (b_k \otimes b_k) d_k$$

$$y_k = D_k d_k + \alpha_k a_k^T d_k a_k + \beta_k b_k^T d_k b_k$$

$$y_k - D_k d_k = \underbrace{\alpha_k a_k^T d_k}_{1} \underbrace{a_k}_{y_k} + \underbrace{\beta_k b_k^T d_k}_{-1} \underbrace{b_k}_{D_k d_k}$$

$\Rightarrow \quad a_k = y_k \qquad b_k = D_k d_k$

SOLVE $\alpha_k (y_k^T d_k) = 1$

$\beta_k ((D_k d_k)^T d_k) = -1$

SO

$$D_{k+1} = D_k + \frac{y_k \otimes y_k}{y_k^T d_k} - \frac{(D_k d_k) \otimes (D_k d_k)}{(D_k d_k)^T d_k} \quad (*)$$

THEOREM 15.2

D_k POS. DEF. X_k GIVEN

IF $t_k > 0$ SELECTED SO THAT

$$X_{k+1} = X_k - t_k \underbrace{D_k^{-1} (\nabla f(X_k))}_{= h_k},$$

SATISFY

3) $\exists\ 0 < \beta < 1$ S.T. $h_k^T \nabla f(X_{k+1}) > \beta h_k^T \nabla f(X_k)$

THEN D_{k+1} DEFINED BY $(*)$ IS POSITIVE DEFINITE

PROOF: 3) $\Rightarrow d_k^T y_k > 0 \quad y_k = A d_k$

$D_k = D_k^{1/2} D_k^{1/2}$ AND THEN FROM DEFINITION

$x^T D_{k+1} x \geq 0$ $\square$

BFGS METHOD

x_k D_k COMPUTED

1, FIND t_k S.T.

$$x_{k+1} = x_k - t_k D_k^{-1} \nabla f(x_k)$$

SATISFY 1, .. 4,

2, SET

$$D_{k+1} = D_k + \frac{y_k \otimes y_k}{d_k^T y_k} - \frac{D_k d_k \otimes D_k d_k}{d_k^T D_k d_k}$$

WHERE $d_k = x_{k+1} - x_k$

$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

NOTE:

NICE THEORETICALLY, AS WELL AS IN PRACTICE.

FOR POS DEF: CHOLESKY FACTORIZATION:

$D_{k+1} = L_{k+1} L_{k+1}^T$ $\quad D_k = L_k L_k^T$

LOWER TRIANGULAR

CLAIM: $\|L_{k+1} - L_k\| \le \|L - L_k\| \; \forall L$ S.T. $LL^T d_k = y_k$