最 適 化 の 手 法

慶大　工学部　柳井　浩

§1　序

　この研究は制約条件なしの最適化問題の解法に於いてよく用いられる conjugate direction method 系の諸手法の体系化の一つの試みである。conjugate direction method はもともと線形連立方程式の一解法として考案されたものであり、その体系化の試みも 2〜3 あるようである。ここでは、できるだけ発見的なやり方で、幾何学的な意味づけを失わぬよう Gram-Schmidt の直交化法からみちびかれる 2つのタイプの共役化過程を中心的視点として conjugate direction method 系の諸手法を見わたすことにした。

　本稿の構成は、まず、§2 に於いて我々のとりあつかう目的関数 $f(x) = \frac{1}{2}x^{T}Ax + b^{T}x + C$ の 2〜3 の性質と；直線上に於ける最小化の反復としての最小化過程；そのうちでも緩和法は、A が対角型の場合、変数の数だけの回数の直線上の最小化の反復で最小点をもたらす事等基本的な事実が示されている。

　次に，§3に於いては，Aを対角型になおすための共役基底の導入とその構成法，A，A^{-1}およびEの共役基底による表現，それを与える漸化式を示した。

　共役基底の構成法のうちでも，本稿で主としてあつかう，Gram-Schmidtの直交化にもとづくものは節をあらため§4でのべた。Gram-Schmidtの直交化にもとづく共役化の方法として，2つのタイプのものが我々の目的に役立ちそうに思われる。Type 1は直交化ののち適当な座標変換を行なうものであり，Type 2はそれまでに共役化されたベクトルにAを乗じたものと直交するベクトルを求めるものである。

　この節ではさらに，これらの過程をベクトルに関する漸化式として書き，その形をもつ一連の方法を例としてあげた。

　§5では，§4の漸化式をさらに行列型に書きあらためると共に，§3に於いて与えたA^{-1}を得る漸化式と組み合わせ，最小化と同時にA^{-1}が得られるアルゴリズムを論じ，その形をもつ一連の方法を例としてあげた。

　§2以下では，文章による説明は少ししか与えていないが，表題，式，図等で充分なように配慮したつもりである。
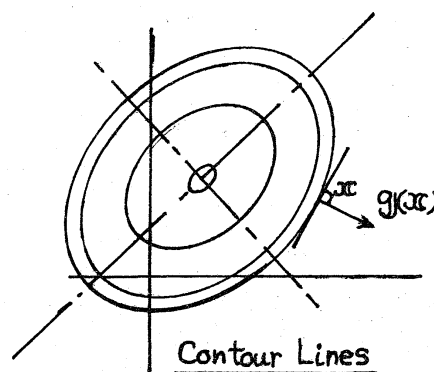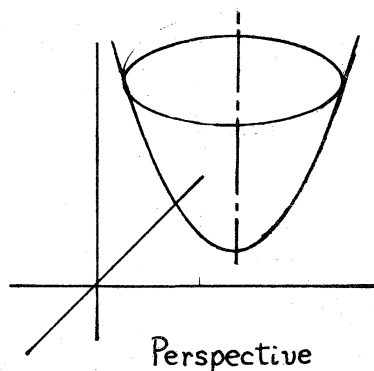
§ 2. Preliminaries

◎ 我々の目的函数 , その勾配 , 最小点, 等に関する性質は次の通りである 。

2.1 Objective Function

$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in R^n, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in R^n, \quad c : real$$

$$\overset{n}{\underset{n}{\Gamma}} A : real, symmetric, positive definite$$

Perspective   Contour Lines

2.2 Gradients

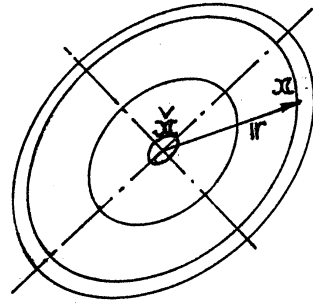$$g(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}, \quad g_* \equiv g(x_*), \quad g(x) = Ax + b$$

2.3 Minimum Point

$$f(x) : Min \ at \ \check{x} \iff g(\check{x}) = 0 \iff A\check{x} + b = 0 \iff \check{x} = -A^{-1}b$$

3

2.4 Residual $r = x - \check{x} = A^{-1} g(x)$

$$r = x - \check{x} = x - (-A^{-1}b) =$$
$$= A^{-1}Ax + A^{-1}b$$
$$= A^{-1}(Ax + b) = A^{-1}g(x)$$

◎ 本稿では、直線上の最小化を単元とする最小化過程を考える。これをめぐる諸式は次の通りである。

2.5 Linear Minimization

Straight Line $x = x_m + \xi t$

base point ⎫
search direction ⎬

Objective Function on $x = x_m + \xi t$

$$\varphi(\xi) = f(x_m + \xi t)$$
$$\varphi'(\xi) = t^T g(x_m + \xi t)$$
$$= t^T(Ax_m + b) + \xi t^T A t$$
$$= t^T g_m + \xi t^T A t$$

Min. Point on $x = x_m + \xi t$

$$x_{m+1} = x_m + \xi_m t$$

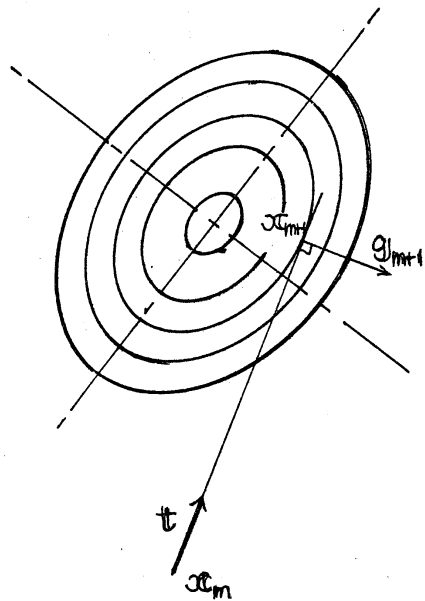min. point $\quad \xi_m = \mu(x_m, t)$

2.6 $\quad \mu(x_m, t) = - \dfrac{t^T g_m}{t^T A t} \quad \Leftarrow \varphi'(\xi_m) = 0$

Lemma 1

2.7 $\quad \boxed{t^T g_{m+1} = 0}$

$$\because \quad 0 = \varphi'(\xi_m) = t^T g(x_m + \xi_m t) = t^T g_{m+1}$$

2.8 Minimization Process

$$\mathcal{X}_{m+1} = \mathcal{X}_m + \mathcal{U}_m$$

$$\mathcal{U}_m \rightleftharpoons \mu(\mathcal{X}_m, t_m)\, t_m = \mathcal{X}_{m+1} - \mathcal{X}_m$$

$$\mathcal{Y}_{i-1} \rightleftharpoons g_i - g_{i-1} = A\,\mathcal{U}_{i-1}$$
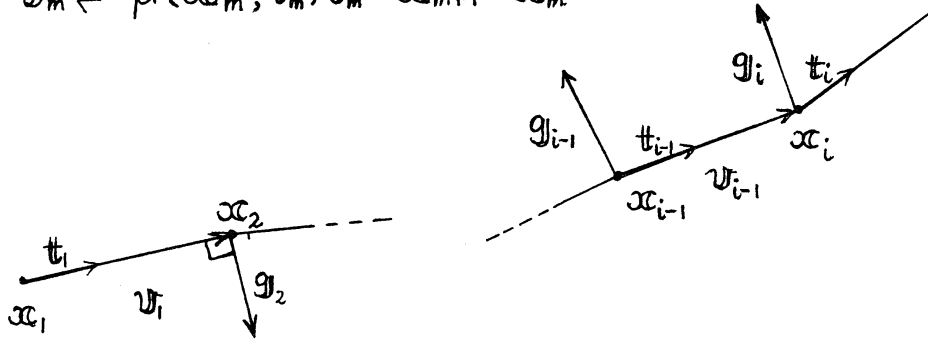
$$\therefore \quad g_i - g_{i-1} = A\mathcal{X}_i + b - (A\mathcal{X}_{i-1} + b) =$$

$$= A(\mathcal{X}_i - \mathcal{X}_{i-1}) =$$

$$= A\,\mathcal{U}_{i-1}$$

2.9 Example (Optimal Gradient Method)

$$t_m = -g_m$$

$$\mathcal{X}_{m+1} = \mathcal{X}_m - \mu(\mathcal{X}_m, -g_m)\, g_m$$
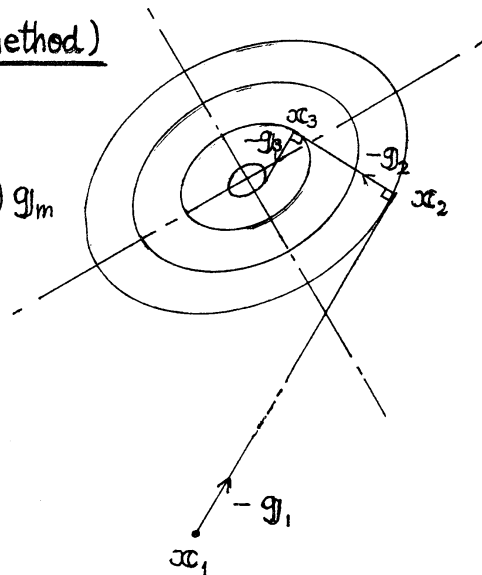
◎最小化過程の最も素朴な例は，次の緩和法であるが，これは関数の形 ── いゝかえれば座標のえらび方で収束性が大中に変る。

2.10 Relaxation Method

$$t_1 = e_1 , \quad t_2 = e_2 , \quad \cdots , \quad t_n = e_n$$

$$t_{n+1} = e_1 , \quad t_{n+2} = e_2 , \quad \cdots , \quad t_{2n} = e_n$$

$$\cdots \qquad \cdots$$

$$x_{m+1} = x_m + \mu (x_m, e.) \, e.$$

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Big\} i$$

a

b

c

d
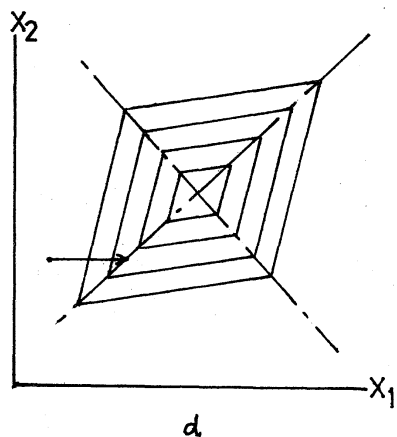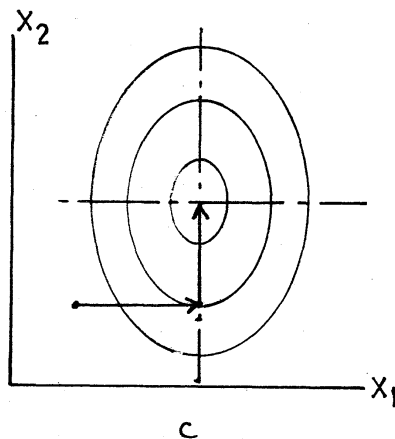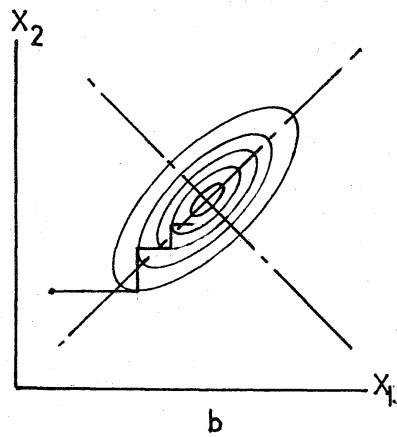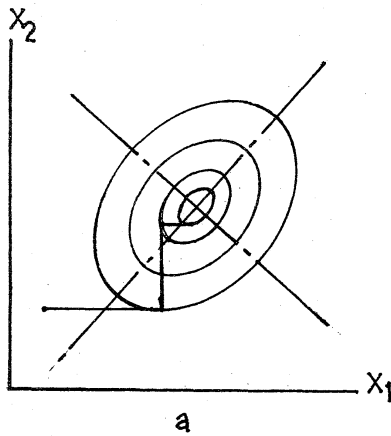
◉ 上図から明らかなように(C)の場合が最も都合がよい。これ

は式でいえば次の通りである。

<u>2.11</u>
<u>Lemma2</u>

$A = \begin{pmatrix} \lambda_1 & & & O \\ & \lambda_2 & & \\ & & \ddots & \\ O & & & \lambda_n \end{pmatrix}$: diagonal, positive definite $(\lambda_1, \lambda_2, \cdots, \lambda_n > 0)$

$\Downarrow$

By relaxation method, x_m reaches $\check{x}$ at most in n linear minimizations, i.e. $m \leq n+1$

<u>Proof</u>

$$f(x) = \tfrac{1}{2}x^T A x + b^T x + c$$

$$= (\tfrac{1}{2}\lambda_1 x_1^2 + b_1 x_1) + (\tfrac{1}{2}\lambda_2 x_2^2 + b_2 x_2) + \cdots + (\tfrac{1}{2}\lambda_n x_n^2 + b_n x_n) + c$$

$\underbrace{\qquad\qquad}$ $\underbrace{\qquad\qquad}$ $\underbrace{\qquad\qquad}$

$\downarrow$
min in
1st min'zation

$\downarrow$
min in
2nd min'zation

$\cdots$

$\downarrow$
min in
n-th min'zation
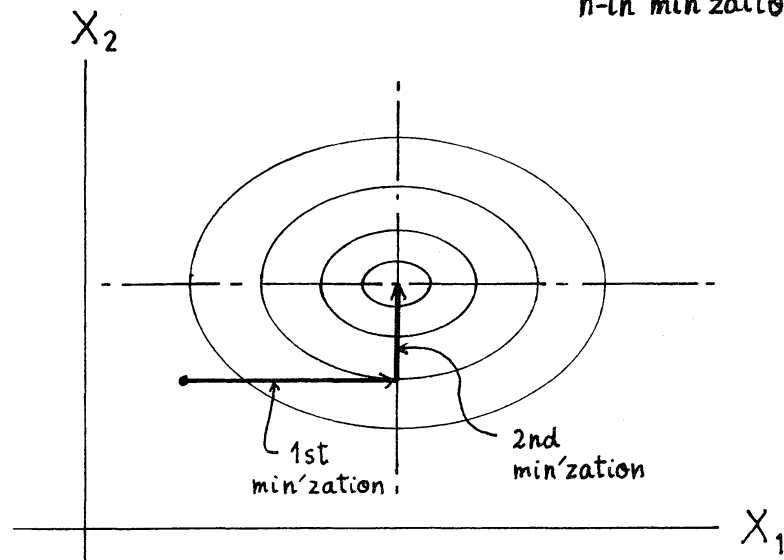
◎ そこで，座標変換によって，A を対角化することが考えられる。2.12 はそのための準備。

2.12 Coordinate Transformations

	OLD COORDINATE	new COORDINATE
BASIS :	$e_1, e_2, \cdots, e_n$	$h_1, h_2, \cdots, h_n$
p :	$p = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n$	$p = z_1 h_1 + z_2 h_2 + \cdots + z_n h_n$
COORDINATE VECTORS :	$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$	$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$

RELATIONS:

$$h_1 = S_{11} e_1 + S_{21} e_2 + \cdots + S_{n1} e_n$$
$$h_2 = S_{12} e_1 + S_{22} e_2 + \cdots + S_{n2} e_n$$
$$\cdots \cdots \cdots$$
$$h_n = S_{1n} e_1 + S_{2n} e_2 + \cdots + S_{nn} e_n$$

$$S_1 = \begin{pmatrix} S_{11} \\ S_{21} \\ \vdots \\ S_{n1} \end{pmatrix}, \quad S_2 = \begin{pmatrix} S_{12} \\ S_{22} \\ \vdots \\ S_{n2} \end{pmatrix}, \cdots, \quad S_n = \begin{pmatrix} S_{1n} \\ S_{2n} \\ \vdots \\ S_{nn} \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$$

$$x = S z, \quad S = (S_1, S_2, \cdots, S_n)$$
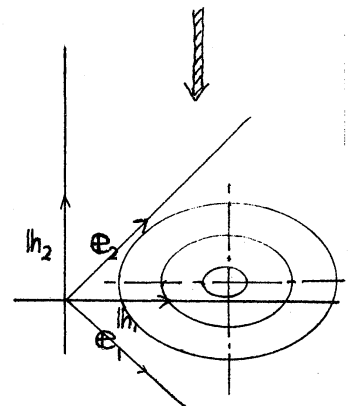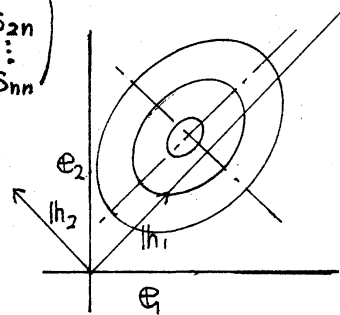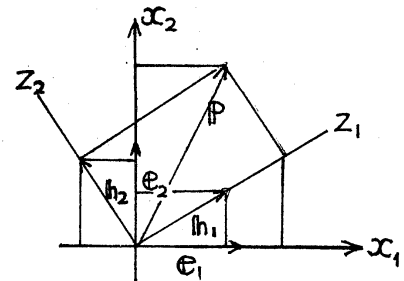
2.13 $f(x)$ in terms of z

$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

$$f_1(z) = \frac{1}{2} z^T S^T A S z^T + (S^T b)^T z + c$$

$$x = S z$$

◉ Ａ の 対角化 の ために は , 互 に 共役 な ベクトル からなる 基

底 , 共役基底 を とればよい 。 ‐ ‐ ‐

§3　Conjugate Directions

3.1　Conjugate Relations

$$\boxed{\begin{array}{l} P : \text{conjugate to } q \\ \quad \text{with respect to } A \\ P, q \in R^n \\ \quad \neq 0 \end{array}} \equiv \boxed{P^T A q = 0}$$

3.2　Canonical Basis

$$\boxed{\{P_1, P_2, \cdots, P_n\} : (A\text{-})\text{canonical basis}} \equiv \boxed{\begin{array}{l} \{P_1, P_2, \cdots, P_n\} : \text{basis} \\ P_i \text{-} P_j \text{ conjugate} \\ \quad i \neq j \end{array}}$$

3.3　Reduction of Quadratic Forms to Canonical Forms

$\{P_1, P_2, \cdots, P_n\}$: canonical basis in terms of a given basis

__Coordinate Transformation__

$$\overset{\text{OLD}}{\downarrow} \quad \overset{\text{NEW}}{\downarrow}$$

$$x = P z \quad , \quad P \equiv (P_1 \vdots P_2 \vdots \cdots \vdots P_n)$$

OLD	NEW
$f(x) = \frac{1}{2} x^T A x + b^T x + c$	$f_1(z) = \frac{1}{2} z^T P^T A P z + (P^T b)^T z + c$

$$P^T A P = \begin{pmatrix} -P_1^T- \\ -P_2^T- \\ \vdots \\ -P_n^T- \end{pmatrix} A (P_1 \vdots P_2 \vdots \cdots \vdots P_n)$$

$$= \begin{pmatrix} P_1^T A P_1 & & & \\ & P_2^T A P_2 & & \\ & & \ddots & \\ & & & P_n^T A P_n \end{pmatrix} \quad : \underline{\text{Canonical Form}} \\ \qquad\qquad (\text{Diagonal})$$

$P_i^T A P_j = 0 \quad i \neq j$

* $P_i^T A P_i$: canonical coefficient $\quad i = 1, 2, \cdots, n$

◉ 共役基底をつかえば、A, A^{-1}, E を次のように表現することができる。また、これらを漸化式の形で書いておくと、最小点, と同時に A^{-1} 等を計算するアルゴリズムが構成できる。

3.4 Representation of A, A^{-1} and E in Canonical Basis

a. Lemma 3

$$\boxed{\{\mathbb{P}_1, \mathbb{P}_2, \cdots, \mathbb{P}_n\} : \text{canonical basis}} \Rightarrow \boxed{A = \sum_{j=1}^{n} \frac{A \mathbb{P}_j \mathbb{P}_j^T A}{\mathbb{P}_j^T A \mathbb{P}_j}}$$

Proof 1

* $\forall x \in R^n \quad \exists_1 \{x_1, x_2, \cdots, x_n\}$ such that

$$x = x_1 \mathbb{P}_1 + x_2 \mathbb{P}_2 + \cdots + x_n \mathbb{P}_n$$

* $A x = x_1 A \mathbb{P}_1 + x_2 A \mathbb{P}_2 + \cdots + x_n A \mathbb{P}_n = \sum_{j=1}^{n} x_j A \mathbb{P}_j$

* $\sum_{j=1}^{n} \frac{A \mathbb{P}_j \mathbb{P}_j^T A}{\mathbb{P}_j^T A \mathbb{P}_j} x = \sum_{j=1}^{n} \frac{A \mathbb{P}_j \mathbb{P}_j^T A}{\mathbb{P}_j^T A \mathbb{P}_j} \sum_{k=1}^{n} x_k \mathbb{P}_k = \sum_{j=1}^{n} x_j A \mathbb{P}_j = A x$

Proof 2 (Heuristics)

(i) $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ given

$\|x\| = 1$ (without loss of generality)

$y = M x, \quad M = \begin{pmatrix} \cdots m_1 \cdots \\ \cdots m_2 \cdots \\ \vdots \\ \cdots m_3 \cdots \end{pmatrix}$

$m_i = \frac{y_i}{z^T x} z^T \qquad z^T x \neq 0$

$\|z\| = 1$ (without loss of generality)

$M = \frac{y z^T}{z^T x}$
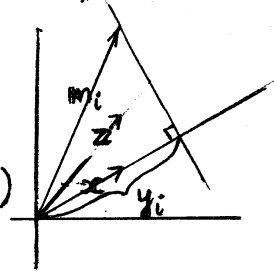
(ii) $\underbrace{A_i \mathbb{P}_i}_{y} = \underbrace{A_i \mathbb{P}_i}_{x} \qquad z = A \mathbb{P}_i$

$A_i = \frac{A \mathbb{P}_i \mathbb{P}_i^T A}{\mathbb{P}_i^T A \mathbb{P}_i}$ no effects on $[A \mathbb{P}_i]^\perp$

$\therefore A = \sum_{i=1}^{n} \frac{A \mathbb{P}_i \mathbb{P}_i^T A}{\mathbb{P}_i^T A \mathbb{P}_i}$

b. Corollary 3.1

$$A_i = \sum_{j=1}^{i} \frac{A P_j P_j^T A}{P_j^T A P_j} \quad \Rightarrow \quad A P_i = A_i P_i$$

c. Corollary 3.2

$$A_1 = \frac{A P_1 P_1^T A}{P_1^T A P_1}$$

$$A_i = A_{i-1} + \frac{A P_i P_i^T A}{P_i^T A P_i} \quad \Rightarrow \quad A_n = A$$

d. Corollary 3.3

$$y_i = A P_i \quad \Rightarrow \quad A^{-1} = \sum_{j=1}^{n} \frac{A^{-1} y_j y_j^T A^{-1}}{y_j^T A^{-1} y_j} = \sum_{j=1}^{n} \frac{P_j P_j^T}{P_j^T A P_j}$$

$$\because \quad y_i^T A^{-1} y_j = P_i^T A A^{-1} A P_j = P_i^T A P_j = 0$$

e. Corollary 3.4

$$E = \sum_{j=1}^{n} \frac{P_j P_j^T A}{P_j^T A P_j} = \sum_{j=1}^{n} \frac{A P_j P_j^T}{P_j^T A P_j}$$

$$= \sum_{j=1}^{n} \frac{y_j y_j^T A^{-1}}{y_j^T A^{-1} y_j} = \sum_{j=1}^{n} \frac{A^{-1} y_j y_j^T}{y_j^T A^{-1} y_j}$$

E : unit matrix

◎ そこで，互に共役な方向に直線最小化をくりかえせば，高々 n 回で最小点に到達する。

3.5 Minimization along Conjugate Directions

a. Conjugate Directions Methods

$$x_{m+1} = x_m + v_m$$
$$v_m = \mu(x_m, t_m)\, t_m$$
$$t_1, t_2, \cdots, t_n : \text{canonical basis}$$
$$(v_i A v_j = 0 \quad i \neq j)$$

b. Exactness of Conjugate Directions Methods

By Conjugate Directions Methods, x_m reaches $\check{x}$ at most in n linear minimizations, i.e. $m \leq n+1$

Proof   See Lemma 2. and 3.3 ∎

◎ 次の補助定理は共役方向に進むときに見られる一般的性質。

c. Lemma 4

$$g_{k+1}^T v_\ell = 0 \qquad \begin{array}{l} k = 1, 2, \cdots, n \\ \ell = 1, 2, \cdots, k \end{array}$$

Proof.

$$x_{k+1} = x_{\ell+1} + (x_{\ell+2} - x_{\ell+1}) + \cdots + (x_{k+1} - x_k)$$
$$= x_{\ell+1} + v_{\ell+1} + v_{\ell+2} + \cdots + v_k$$
$$= x_{\ell+1} + \sum_{i=\ell+1}^{k} v_i$$

$$g_{k+1} = A x_{k+1} + b$$
$$= A x_{\ell+1} + b + \sum_{i=\ell+1}^{k} A v_i$$
$$= g_{\ell+1} + \sum_{i=\ell+1}^{k} A v_i$$

$$g_{k+1}^T v_\ell = \underbrace{g_{\ell+1}^T v_\ell}_{\overset{\shortparallel}{0} \text{ (Lemma 1)}} + \underbrace{\sum_{i=\ell+1}^{k} v_i^T A v_\ell}_{\overset{\shortparallel}{0} \text{ (a)}} = 0$$
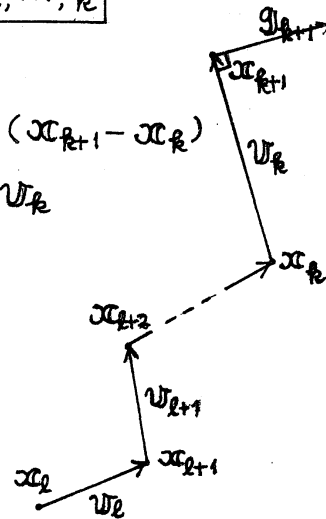
◎ 共役基底を構成するにはいくつかの方法がある。

3.6 Construction of Canonical Basis

a. Complete Squares

Ex.

$$(x_1, x_2) \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= a_{11} x_1^2 + 2 a_{12} x_1 x_2 + a_{22} x_2^2$$

$$= a_{11} x_1^2 - \frac{a_{12}^2}{a_{22}} x_1^2 + a_{22} \left(\frac{a_{12}^2}{a_{22}^2} x_1^2 + 2 \frac{a_{12}}{a_{22}} x_1 x_2 + x_2^2 \right)$$

$$= \frac{a_{11} a_{22} - a_{12}^2}{a_{22}} x_1^2 + a_{22} \left(\frac{a_{12}}{a_{22}} x_1 + x_2 \right)^2$$

$$= \frac{a_{11} a_{22} - a_{12}^2}{a_{22}} z_1^2 + a_{22} z_2^2$$

$$z_1 \equiv x_1$$

$$z_2 \equiv \frac{a_{12}}{a_{22}} x_1 + x_2$$

b. Jacobi's Method

Ex.
Canonical Basis.
$$\begin{cases} p_1 = e_1 \\ p_2 = \alpha_1^1 e_1 + e_2 \\ p_3 = \alpha_1^2 e_1 + \alpha_2^2 e_2 + e_3 \end{cases}$$

e_1, e_2, e_3 : given basis

$$p_2^T A e_1 = 0 \qquad \alpha_1^1 e_1^T A e_1 + e_2^T A e_2 = 0$$

$$p_3^T A e_1 = 0 \qquad \alpha_1^2 e_1^T A e_1 + \alpha_2^2 e_2^T A e_1 + e_3^T A e_1 = 0$$

$$p_3^T A e_2 = 0 \qquad \alpha_1^2 e_1^T A e_2 + \alpha_2^2 e_2^T A e_2 + e_3^T A e_2 = 0$$

13

c. Middle Points of Chords

$$\mathbb{p}^T A \mathbb{q} = 0$$

$$x^T A x = const$$
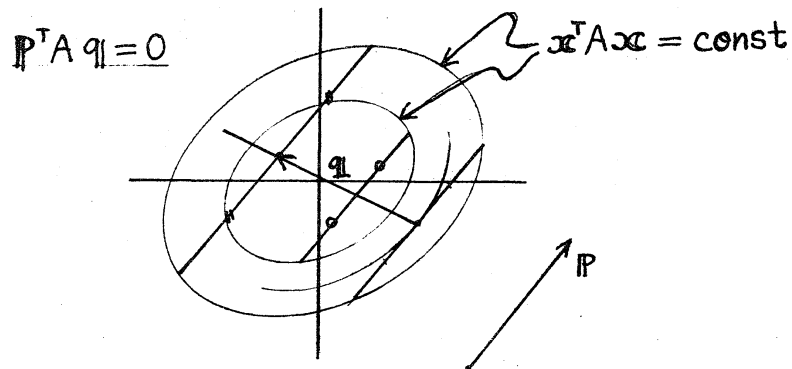
◎ 次 の 2 つ は ， 最適化 に よ く 用 い ら れ る も の だ が ， 本橋 で は

c を 中心 に 考 え る 。 C に つ い て は §4 で く わ し く の べ る 。

d. Two Sub-Optimal Points
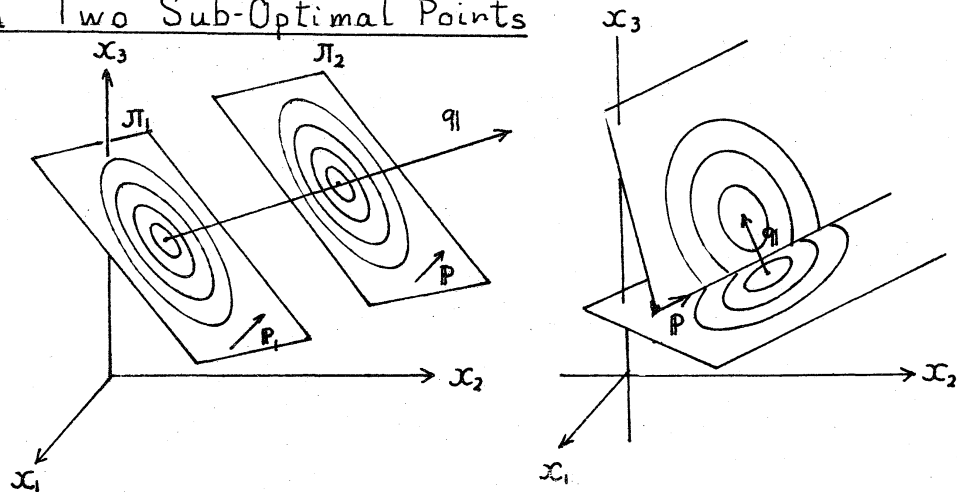
$$\mathbb{p}^T A \mathbb{q} = 0$$

cf. (PARallel TANgent Algorithm
Powell-Zangwill's Algorithm

e. Gram Schmidt Orthogonalization Process

cf. (Fletcher Reeves's(Polak-Ribière) Algorithm
Davidon Fletcher Powell's Algorithm

◎ 以下では，Gram Schmidt 法にもとづく共役化を考える。

§4 Recurrence Relations in Vector Form

4.1 Gram Schmidt Orthogonalization

a.

$$\mathbb{P}_1^*, \mathbb{P}_2^*, \cdots, \mathbb{P}_n^* \quad \text{given basis}$$
$$\mathbb{t}_1^* \fallingdotseq \mathbb{P}_1^*$$
$$\mathbb{t}_i^* \fallingdotseq \mathbb{P}_i^* - \sum_{j=1}^{i-1} \frac{\mathbb{P}_i^{*T}\mathbb{t}_j^*}{\mathbb{t}_j^{*T}\mathbb{t}_j^*}\mathbb{t}_j^* \quad i=1,\cdots,n$$

$$\Rightarrow \quad \{\mathbb{t}_1^*, \mathbb{t}_2^*, \cdots, \mathbb{t}_n^*\}: \text{orthogonal basis}$$

◎ Type 1 は直交基底を座標変換によって共役基底になおす。

b. Construction of Canonical Basis — Type 1

(i) * A : positive definite $\Rightarrow$ $^{\exists}J$ such that

$$^n_n J : \text{regular}, \quad J^T J = A$$

* $\mathbb{P} - \mathbb{q}$: conjugate $\iff$ $\mathbb{p}^* = J\mathbb{p}, \mathbb{q}^* = J\mathbb{q}$: orthogonal

$$\because \mathbb{p}^T A \mathbb{q} = \mathbb{p}^T J^T J \mathbb{q} = (J\mathbb{p})^T(J\mathbb{q}) = \mathbb{p}^{*T}\mathbb{q}^*$$

(ii)

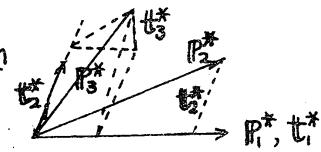| orthogonal basis	canonical basis
$\{\mathbb{P}_1^*, \mathbb{P}_2^*, \cdots, \mathbb{P}_n^*\}$: given basis	$\{\mathbb{P}_1, \mathbb{P}_2, \cdots, \mathbb{P}_n\}$: given basis
	$\mathbb{P}_i^* = J\mathbb{P}_i$
$\mathbb{t}_1^* \fallingdotseq \mathbb{P}_1^*$	$J\mathbb{t}_1 = J\mathbb{P}_1$
$\mathbb{t}_i^* \fallingdotseq \mathbb{P}_i^* - \sum_{j=1}^{i-1}\frac{\mathbb{P}_i^{*T}\mathbb{t}_j^*}{\mathbb{t}_j^{*T}\mathbb{t}_j^*}\mathbb{t}_j^*$	$J\mathbb{t}_i = J\mathbb{P}_i - \sum_{j=1}^{i-1}\frac{\mathbb{P}_i^T J^T J \mathbb{t}_i}{\mathbb{t}_j^T J^T J \mathbb{t}_j}J\mathbb{t}_j$ $\quad\overset{=A}{\frown}$
	$\xrightarrow{J^{-1}}$

$$\boxed{\begin{array}{l}\mathbb{t}_1 = \mathbb{P}_1 \\ \mathbb{t}_i = \mathbb{P}_i - \sum_{j=1}^{i-1}\frac{\mathbb{P}_i^T A \mathbb{t}_j}{\mathbb{t}_j^T A \mathbb{t}_j}\mathbb{t}_j \\ i=1,2,\cdots,n\end{array}}$$

| $\{\mathbb{t}_1^*, \mathbb{t}_2^*, \cdots, \mathbb{t}_n^*\}$ orthogonal basis | $\{\mathbb{t}_1, \mathbb{t}_2, \cdots, \mathbb{t}_n\}$ canonical basis |

◎ Type 2 は, それまでに求めた共役系 $P_1 \cdots P_{i-1}$ に A を乗じた系 $AP_1, \cdots, AP_{i-1}$ と直交するベクトルを Gram-Schmidt 法で作る。

c. Construction of Canonical Basis —— Type 2

$\{P_1, P_2, \cdots, P_n\}$: given basis

$$t_1 = P_1$$
$$u_1 = At_1$$
$$u_j = At_j - \sum_{\ell=1}^{j-1} \frac{u_\ell^T A t_j}{u_\ell^T u_\ell}$$
$$t_i = P_i - \sum_{j=1}^{i-1} \frac{P_i^T u_j}{u_j^T u_j} u_j$$
$$i = 1, 2, \cdots, n$$

orthogonal basis
for $[At_1, At_2, \cdots, At_{j-1}]$

vector orthogonal to
$[At_1, At_2, \cdots At_{i-1}]$

$\{t_1, t_2, \cdots, t_n\}$ canonical basis

◉ ところで, 最初に与える「任意の基底」として, 実際につかえそうなものをしらべておく。

4.2 Practical Selection of the "Given Basis"

a. $P_i \equiv e_i$

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

b. $\mathbb{P}_i \fallingdotseq g_i$ (so far as $g_i \neq 0$)

(b' $\mathbb{P}_i \fallingdotseq y_{i-1} = g_i - g_{i-1}$, $\mathbb{P}_1 = g_1$)

$\underline{\text{Remark 1}}$ $g_i \notin [t_1, t_2, \cdots, t_{i-1}]$

$$\because \begin{cases} g_i \neq 0 \\ g_i^T t_j = 0 , \quad j = 1, \cdots, i-1 \end{cases}$$

$\underline{\text{Remark 2}}$ $g_i \notin [At_1, At_2, \cdots, At_{i-1}]$

$\because$ (i) $g_i \in [At_1, At_2, \cdots, At_{i-1}]$

$\Rightarrow \exists (\alpha_1, \alpha_2, \cdots, \alpha_{i-1})$ such that

$$g_i = \alpha_1 At_1 + \cdots + \alpha_{i-1} At_{i-1}$$

(ii) $t_j^T g_i = 0$, $j = 1, 2, \cdots, i-1$

$$\because \begin{cases} t_j \,/\!/\, v_j \\ v_j^T g_i = 0 \quad j = 1, 2, \cdots, j \quad (\text{Lemma 4}) \end{cases}$$

(iii) $\alpha_j = 0$ $j = 1, 2, \cdots, i-1$

$$\because \quad 0 \underset{\substack{\uparrow \\ \text{(ii)}}}{=} t_j^T g_i \underset{\substack{\uparrow \\ \text{(i)}}}{=} \alpha_1 \underbrace{t_j^T A t_1}_{\substack{\| \\ 0}} + \cdots + \alpha_j \underbrace{t_j^T A t_j}_{\substack{\neq \\ 0}} + \cdots + \alpha_{i-1} \underbrace{t_j^T A t_{i-1}}_{\substack{\| \\ 0}}$$

(iv) $\left.\begin{array}{c}\text{(i)} \\ \text{(iii)}\end{array}\right\} \Rightarrow g_i = 0$ ∴ contradiction

c. $\mathbb{P}_i \fallingdotseq r_i$ (so far as $r_i \neq 0$)

r_i : residual

◎ Type 1 の共役化をすれば最小化の方向は次のようになる。

a' は a の変形である。

4.3 Minimization Algorithms deduced from Type 1 Construction

a. $P_1, P_2, \cdots, P_n$ given basis

$$t_1 = P_1$$
$$t_i = P_i - \sum_{j=1}^{i-1} \frac{P_i^T A t_j}{t_j^T A t_j} t_j$$

a' $P_1, P_2, \cdots, P_n$ given basis

$$t_1 = P_1$$
$$t_i = P_i + \sum_{j=1}^{i-1} \frac{P_i^T (g_{j+1} - g_j)}{t_j^T g_j} t_j \left(= P_i + \sum_{j=1}^{i-1} \frac{P_i^T y_j}{t_j^T g_j} t_j \right)$$

$$\because A t_j // A v_j = A(x_{j+1} - x_j)$$
$$= A x_{j+1} + b - (A x_j + b)$$
$$= g_{j+1} - g_j = y_j$$

$$\frac{P_i^T A t_j}{t_j^T A t_j} = \frac{P_i^T (g_{j+1} - g_j)}{t_j^T (\underbrace{g_{j+1} - g_j}_{\overset{\|}{0}})} = -\frac{P_i^T (g_{j+1} - g_j)}{t_j^T g_j} = -\frac{P_i^T y_j}{t_j^T g_j}$$

◎ この場合には, 最初の基底と勾配の間に直交性がでてくる

b. given basis and gradients

$$g_{i+1}^T P_j = 0 \qquad j = 1, 2, \cdots, i$$

$$\because g_{i+1}^T t_j = 0 \quad j = 1, 2, \cdots, i \quad (\text{Lemma 4})$$
$$0 = g_{i+1}^T \left(P_j + \sum_{k=1}^{i-1} \frac{P_j^T y_j}{t_j^T g_j} t_j \right) = g_{i+1}^T P_j$$

◉ この漸化式をさらに簡単にするために，1階の漸化式になるための条件を求める。

c. Reduction to the Recurrence Relations of the 1-st Order

$$\boxed{\begin{array}{l} P_i^T A t_j = 0 \\ j = 1, 2, \cdots, i-2 \end{array}} \quad \Rightarrow \quad \boxed{\; t_i = P_i + \frac{P_i^T y_{i-1}}{P_{i-1}^T g_{i-1}} t_{i-1} \;}$$

Proof

$$t_i = P_i - \sum_{j=1}^{i-1} \frac{P_i^T A t_j}{t_j^T A t_j} t_j = P_i - \frac{P_i^T A t_{i-1}}{t_{i-1}^T A t_{i-1}} t_{i-1} =$$

$$= P_i - \frac{P_i^T (g_i - g_{i-1})}{\left(P_{i-1} - \frac{P_{i-1}^T A t_{i-2}}{t_{i-2}^T A t_{i-2}} t_{i-2} \right)^T (g_i - g_{i-1})} t_{i-1} =$$

$$= P_i - \frac{P_i^T (g_i - g_{i-1})}{P_{i-1}^T g_i - P_{i-1}^T g_{i-1}} t_{i-1} = \qquad (Lemma\ 4)$$

$$= P_i + \frac{P_i^T y_{i-1}}{P_{i-1}^T g_{i-1}} t_{i-1} \qquad ▨$$

◉ 上のような条件をみたす基底として $-g_i$ が考えられる。

Fletcher Reeves の方法はこうしてみちびかれる。

Example

$$P_i = -g_i \quad (so\ far\ as\ g_i \neq 0)$$

$$\ast \quad \left. \begin{array}{l} P_i^T A t_j = 0 \\ g_i^T g_j = 0 \end{array} \right\} \quad j = 1, 2, \cdots, i-1$$

$$P_i^T g_j = 0 \quad j = 1, 2, \cdots, i-1 \Rightarrow g_i^T g_j = 0 \quad j = 1, 2, \cdots, i-1$$

$$P_i^T A t_j \propto g_i^T (g_{j+1} - g_j) = \begin{array}{l} + g_i^T g_{j+1} \; \text{---} =0 \quad j = 1, \cdots, i-2 \\ - g_i^T g_j \; \text{---} =0 \quad j = 1, \cdots, i-1 \\ = 0 \end{array}$$

$$\boxed{\begin{array}{l} t_i = -g_i + \dfrac{g_i^T (g_i - g_{i-1})}{g_{i-1}^T g_{i-1}} t_{i-1} \quad \text{--- Polak Rivière's Algorithm} \\[2mm] \quad = -g_i + \dfrac{g_i^T g_i}{g_{i-1}^T g_{i-1}} t_{i-1} \quad \text{--- Fletcher Reeves's Algorithm} \end{array}}$$

⊙ 一般に，k 階の漸化式に帰着する条件も求められる。

d. Reduction to the Recurrence Relations of the k-th Order

$$
\boxed{\begin{array}{l} \mathbb{P}_i^T A \, \mathfrak{t}_j = 0 \\ j = 1, 2, \cdots, i-k-1 \end{array}} \quad \Rightarrow \quad \boxed{\mathfrak{t}_i = \mathbb{P}_i + \sum_{j=i-k}^{i-1} \frac{\mathbb{P}_i^T y_j}{\mathbb{P}_j^T y_j} \mathfrak{t}_j}
$$

Proof.　abbr.　cf. c.　■

⊙ 上のような条件をみたす $\mathbb{P}_i$ も無くはない。(次の例題の数値的な有効性はまだ調べていない。)

Example

$$\mathbb{P}_i = -y_{i-1} = -(g_i - g_{i-1}), \quad i = 2, \cdots, n, \quad \mathbb{P}_1 = -g_1$$

$$* \quad \mathbb{P}_i^T A \, \mathfrak{t}_j = 0 \qquad j = 1, 2, \cdots, i-2$$

$$\because \quad \mathbb{P}_j^T g_i = 0 \qquad j = 1, 2, \cdots, i-1$$

$$\Rightarrow \begin{cases} g_1^T g_i = 0 \\ (g_j - g_{j-1})^T g_i = 0 \qquad j = 2, \cdots, i-1 \end{cases}$$

$$\mathbb{P}_i A \, \mathfrak{t}_j \propto (g_i - g_{i-1})^T (g_{j+1} - g_j) =$$

$$\qquad = g_i^T (g_{j+1} - g_j) \qquad ----- 0 \qquad j = 1, 2, \cdots, i-2$$

$$\qquad - g_{i-1}^T (g_{j+1} - g_j) \qquad ----- 0 \qquad j = 1, 2, \cdots, i-3$$

$$\qquad = 0 \qquad j = 1, 2, \cdots, i-3, \quad i \geqq 2$$

$$* \quad \boxed{\mathfrak{t}_i = -y_{i-1} + \frac{y_{i-1}^T y_{i-1}}{y_{i-2}^T g_{i-1}} \mathfrak{t}_{i-1} + \frac{y_{i-1}^T y_{i-2}}{y_{i-3}^T g_{i-2}} \mathfrak{t}_{i-2}}$$

⊙ Type 2 の方からもっと簡単で実用的なアルゴリズムが出来るか否かは未調査である。

◉ §4 の漸化式を行列型に書き改めてみよう。

§5 Recurrence Relations in Matrix Form

◉ まず、Type 1 の場合については、

5.1. Recurrence Relations deduced from Type 1 Construction

a.

Vector Form	Matrix Form
$t_1 = p_1$	$t_1 = E p_1$
$t_i = p_i - \sum_{j=1}^{i-1} \dfrac{t_j^T A p_i}{t_j^T A t_j} t_j$	$t_i = \left(E - \sum_{j=1}^{i-1} \dfrac{t_j t_j^T A}{t_j^T A t_j} \right) p_i$
$\left(= p_i + \sum_{j=1}^{i-1} \dfrac{y_j^T p_i}{t_j^T g_j} t_j \right)$	$\left(= \left(E + \sum_{j=1}^{i-1} \dfrac{t_j y_j^T}{t_j^T g_j} \right) p_i \right)$

$$\because \quad t_i = p_i - \sum_{j=1}^{i-1} \frac{t_j^T A p_i}{t_j^T A t_j} t_j = p_i - \sum_{j=1}^{i-1} t_j \frac{t_j^T A p_i}{t_j^T A t_j} = \left(E - \sum_{j=1}^{i-1} \frac{t_j t_j^T A}{t_j^T A t_j} \right) p_i$$

b. Recurrence Relations in Matrix Form

$$C_i' \equiv E - \sum_{j=1}^{i-1} \frac{t_j t_j^T A}{t_j^T A t_j} \left(= E + \sum_{j=1}^{i-1} \frac{t_j y_j^T}{t_j^T g_j} \right)$$

$$\boxed{\begin{array}{l} t_i = C_i' p_i \\ C_i' = C_{i-1}^1 - \dfrac{t_{i-1} t_{i-1}^T A}{t_{i-1}^T A t_{i-1}} \left(= C_{i-1}^1 + \dfrac{t_{i-1} y_{i-1}^T}{t_{i-1}^T g_{i-1}} \right) \end{array}}$$

Remark 1

C_1' needs not be equal to E, but it is necessary that $C_1 p_1 \notin [p_2, p_3, \cdots, p_n]$. In this case, $C_1 p_1$ is to be regarded as the first member of the given basis.

Remark 2

Since $\alpha\, C_i^1$ $(i=1,2,\cdots,n \quad \alpha\neq 0)$ give the same search directions as C_i^1, we may replace C_i^1 by $\widetilde{C}_i^1 \equiv \alpha\, C_i^1$.

In this case, the recurrence relation turns out to be

$$\widetilde{C}_i^{\;1} = \widetilde{C}_{i-1}^{\;1} - \alpha\, \frac{t_{i-1}\, t_{i-1}^T A}{t_{i-1}^T A\, t_{i-1}}$$

Remark 3

$$C_n^1 = 0$$

See 3.4 e. Cor 3.4

Remark 4

$$C_i^1 \quad (i=1,2,\cdots,n)\; :\; idempotent$$

$$\because\; * \; P_{i-1} \equiv \sum_{j=1}^{i-1} \frac{t_j\, t_j^T A}{t_j^T A\, t_j} \quad :\; conjugate\ projection\ matrix$$

$$\Downarrow$$

$$idempotent$$

$$c.f. \quad \frac{t_j\, \overbrace{t_j^T A \cdot t_j}\, t_j^T A}{t_j^T A\, t_j \quad t_j^T A\, t_j} = \frac{t_j\, t_j^T A}{t_j^T A\, t_j}$$

$$\frac{t_i\, \overbrace{t_i^T A \cdot t_j}\, t_j^T A}{t_i^T A\, t_i \quad t_j^T A\, t_j} = 0 \qquad i\neq j$$

$$* \; C_i^1 = E - P_{i-1}\; :\; idempotent$$

Remark 5

$$C_j^1\, r_i = r_i \qquad j = 1,\cdots,i-1$$

$$(\, r_i = A^{-1} g_i\; :\; residual\,)$$

$$\because\; C_{i-1}^1\, r_i = C_{i-1}^1\, A^{-1} g_i = C_{i-2}^1\, A^{-1} g_i - \overbrace{\frac{t_{i-2}\, t_{i-2}^T A A^{-1} g_i}{t_{i-2}^T A\, t_{i-2}}}^{\substack{0 \\ \| \quad (Lemma\ 4)}}$$

$$= C_{i-2}^1\, r_i = \cdots = C_1^1\, r_i = r_i$$

◎ Type 2 の 場合 にも 同 様 に して ,

5.2 Recurrence Relation deduced from Type 2 Construction

a.

Vector Form	Matrix Form
$t_1 = p_1$	$t_1 = Ep_1$
$u_1 = At_1$	$u_1 = At_1$
$u_j = At_j - \sum_{\ell=1}^{j-1} \dfrac{u_\ell^T At_j}{u_\ell^T u_\ell} u_\ell$	$u_j = \left(E - \sum_{\ell=1}^{j-1} \dfrac{u_\ell u_\ell^T}{u_\ell^T u_\ell}\right) At_j$
$t_i = p_i - \sum_{j=1}^{i-1} \dfrac{p_i^T u_j}{u_j^T u_j} u_j$	$t_i = \left(E - \sum_{j=1}^{i-1} \dfrac{u_j u_j^T}{u_j^T u_j}\right) p_i$
$i = 1, 2, \cdots, n$	$i = 1, 2, \cdots, n$

$$\because \quad u_j = At_j - \sum_{\ell=1}^{j-1} \frac{u_\ell^T At_j}{u_\ell^T u_\ell} u_\ell$$

$$= At_j - \sum_{\ell=1}^{j-1} u_\ell \frac{u_\ell^T At_j}{u_\ell^T u_\ell}$$

$$= \left(E - \sum_{\ell=1}^{j-1} \frac{u_\ell u_\ell^T}{u_\ell^T u_\ell}\right) At_j$$

$$t_i = p_i - \sum_{j=1}^{i-1} \frac{p_i^T u_j}{u_j^T u_j} u_j$$

$$= p_i - \sum_{j=1}^{i-1} u_j \frac{u_j^T p_i}{u_j^T u_j}$$

$$= \left(E - \sum_{j=1}^{i-1} \frac{u_j u_j^T}{u_j^T u_j}\right) p_i$$

b. Recurrence Relation in Matrix Form

$$C_i^2 \equiv \left(E - \sum_{j=1}^{i-1} \frac{u_j u_j^T}{u_j^T u_j}\right)$$

$$\boxed{\begin{aligned} t_1 &= C_1^2 \, p_1 \\ C_i^2 &= C_{i-1}^2 - \frac{C_{i-1}^2 \, y_{i-1} \, y_{i-1}^T \, C_{i-1}^{2T}}{(C_{i-1}^2 y_{i-1})^T (C_{i-1}^2 y_{i-1})} \end{aligned}}$$

$$\because \quad C_i^2 = C_{i-1}^2 - \frac{u_{i-1} u_{i-1}^T}{u_{i-1}^T u_{i-1}} \quad , \quad u_{i-1} = C_{i-1}^2 A t_{i-1}$$

$$= C_{i-1}^2 - \frac{C_{i-1}^2 A t_{i-1} t_{i-1}^T A C_{i-1}^{2T}}{t_{i-1}^T A C_{i-1}^{2T} C_{i-1}^2 A t_{i-1}}$$

$$= C_{i-1}^2 - \frac{C_{i-1}^2 \, y_{i-1} \, y_{i-1}^T \, C_{i-1}^{2T}}{(C_{i-1}^2 y_{i-1})^T (C_{i-1}^2 y_{i-1})}$$

23

Remark 1

C_1^2 needs not be equal to E. cf. Remark 1 in 5.1,b.

Remark 2

$\tilde{C}_i^2$ may be replaced by $\tilde{C}_i^2 \eqcirc \alpha\, C_i^2$, which is given

by the recurrence relation

$$\tilde{C}_i^2 = \tilde{C}_{i-1}^2 - \alpha\, \frac{C_{i-1}^2\, y_{i-1}\, y_{i-1}^T\, C_{i-1}^{2\,T}}{(C_{i-1}\, y_{i-1})^T (C_{i-1}\, y_{i-1})}$$

cf. Remark 2 in 5.1 b

Remark 3

$$C_h^2 = 0$$

see 3.4 e. Cor. 3.4

Remark 4

C_i^2 : symmetric, idempotent

(orthogonal projection to $[At_1, \cdots, At_{i-1}] = [y_1, \cdots, y_{i-1}]$)

Remark 5

$$C_i^2\, y_\ell = y_\ell \qquad \ell \leqq i-1$$

$$\because \quad C_i^2\, y_\ell = \left(E - \sum_{j=1}^{i-1} \frac{u_j\, u_j^T}{u_j^T u_j}\right) y_\ell$$

$$= y_\ell \qquad u_j \in [At_1, \cdots, At_{i-1}] \perp t_i, t_{i+1}, \cdots$$

◉　最小点と共に，A^{-1}も求めたい場合には，

5.3　Construction of A^{-1}

* $A^{-1} = \sum_{i=1}^{n} \dfrac{t_i t_i^T}{t_i^T A t_i}$ 　　　　 $(3.4, d. Cor. 3.3)$

　　　　　 $= \sum_{i=1}^{n} \dfrac{v_i v_i^T}{v_i^T A v_i}$ 　　　 $\because t_i \parallel v_i$

* $D_i \equiv \sum_{j=1}^{i-1} \dfrac{t_j t_j^T}{t_j^T A t_j} = \sum_{j=1}^{i-1} \dfrac{v_j v_j^T}{v_j^T A v_j}$ 　 $\Rightarrow$ 　 $D_{n+1} = A^{-1}$

$$
\boxed{\begin{array}{l} D_1 = 0 \\[4pt] D_i = D_{i-1} + \dfrac{t_{i-1} t_{i-1}^T}{t_{i-1}^T A t_{i-1}} \end{array}}
$$
　 $\Rightarrow$ 　 $D_{n+1} = A^{-1}$

* D_i : symmetric

* $D_i g_k = 0$ 　　 $k \geq i$

　　 $\because D_i g_k = \sum_{j=1}^{i-1} \dfrac{v_j \overbrace{v_j^T g_k}^{0}}{v_j^T A v_j} = 0$

◎ さらに，A^{-1} を もとめる漸化式と最小値を求める漸化式をまとめて，1つの漸化式を作ることもできる。

5.4 Combined Algorithms

__Example__ (Combined Algorithm, Davidon-Fletcher-Powell)

$$
\boxed{
\begin{aligned}
P_i &= g_i \\
H_i^2 &= C_i^2 + D_i \\
t_i &= H_i^2 P_i
\end{aligned}
}
$$

* $H_i^2 g_k = C_i^2 g_k \qquad k \geq i$

     $\because \; H_i g_i = (C_i^2 + D_i) g_k$

$$
= C_i^2 g_k + \sum_{j=1}^{i-1} \frac{t_j \, t_j^T}{t_j^T A \, t_j} \, g_k
$$

$$
= C_i^2 g_k
$$

* $H_i^2 y_k = C_i^2 y_k \qquad k \geq i$

* H_i^2 symmetric

$$
H_i^2 = H_{i-1}^2 - \frac{H_{i-1}^2 \, y_{i-1} \, y_{i-1}^T H_{i-1}^2}{(H_{i-1}^2 y_{i-1})^T (H_{i-1}^2 y_{i-1})} + \frac{v_{i-1} \, v_{i-1}^T}{v_{i-1}^T A \, v_{i-1}}
$$

__Example__ (Combined Algorithm)

$$
\boxed{
\begin{aligned}
P_i &= g_i \\
H_i^1 &= C_i^1 + D_i \\
t_i &= H_i^1 P_i
\end{aligned}
}
$$