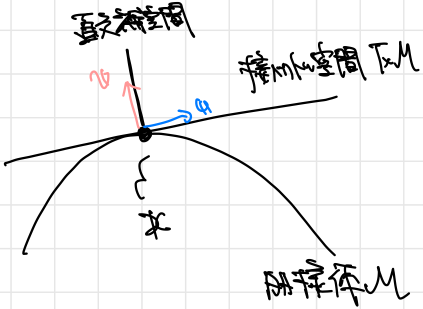
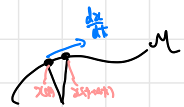


多様体上を動く点をパラメータtを使って$x(t)$と表現する.

接ベクトル空間の元は $u = \frac{dx}{dt}$ と表現される.

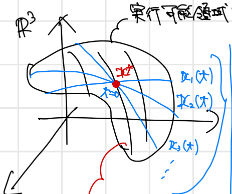


$\boxed{\text{考え方}}$

実行可能領域上をパラメータtによって動く点を$x(t)$とおく.

$x(t=0) = x^*$ とする.

実行可能領域 $S=\{x \mid g_i(x)=0,\ i=1,2,\cdots,m\}$



1次元のパラメータtを動かすことで
実行可能領域の中でx^*を通る
曲線$x_i(t)$を定義する.
無限本の$x(t)$
任意の$x(t)$に対してx^*における
fのt方向の微分が0 $\left(\left.\frac{df}{dt}\right|_{t=0}=0\right)$
になることを確認すればよい.

実行可能領域をm次元のパラメータ $t \in \mathbb{R}^{n-m}$ で
超曲面を表現するのではなく, 1次元方向の移動を
任意な方向に対して行うことで超曲面をカバーする.

接ベクトル空間の例

例: $\|x\|_2^2 = 1$ の接ベクトル空間 T_xM

$x(t)^T x(t) = 1$ より

$$\left(\frac{dx}{dt}\right)^T x(t) + x(t)^T \frac{dx}{dt} = 0 \iff \underbrace{\left(\frac{dx}{dt}\right)^T}_{=u} x = 0$$

$$\therefore\ T_xM : \{u \mid u^T x = 0\}$$

例: シュティーフェル多様体 $X^T X = I$

$$\left(\frac{dX}{dt}\right)^T X + X^T \left(\frac{dX}{dt}\right) = 0 \text{ より.}$$

$$T_xM : \{U \mid U^T X + X^T U\}$$

例6.8: シュティーフェル多様体

$$\mathrm{St}(p,n) = \{X \in \mathbb{R}^{n\times p} \mid X^TX = I_p\} \quad (p \le n)$$

〜 $X \in \mathbb{R}^{n\times p}$ の各列ベクトルが直交している制約

の接ベクトル空間は

$$T_X\mathrm{St}(p,n) = \{Z \in \mathbb{R}^{n\times p} \mid Z^TX + X^TZ = 0\}$$

これは $(X\ X_\perp)$ が直交行列となるような $n\times(n-p)$ 行列 $X_\perp$ を用いて以下のようにも書ける。

$$T_X\mathrm{St}(p,n) = X\mathrm{Skew}_p + X_\perp\mathbb{R}^{(n-p)\times p}$$

最初の表現は $X(t)^TX(t) = I_p$ を t で微分する事によって得られる。

ここで、 X はフルランクであるから $n\times(n-p)$ 行列 $X_\perp$ で

$$X^TX_\perp = 0, \quad X_\perp^TX_\perp = I_{n-p}$$

$$n\Big\{\,[\,\overset{p}{X}\quad \overset{n-p}{X_\perp}\,]$$

$X_\perp$ の列の直交補空間の基底を列にもつ行列

となるようなもの、すなわち $X^TX = I_p$ と合わせて $(X\ X_\perp)$ が直交行列となるようなものが存在する。よって、適当な $n\times p$ 行列 P が存在して

$$\underbrace{\dot{X}}_{=Z} = (X\ X_\perp)P$$

接ベクトル空間の元 Z を $Z = [X\ X_\perp]\underbrace{\begin{bmatrix}\Omega\\ K\end{bmatrix}}_{P}$ で表現

と表せる。ここで P を $p\times p$ 行列 Ω と $(n-p)\times p$ 行列 K によってブロック化する事で

$$\dot{X} = (X\ X_\perp)\begin{pmatrix}\Omega\\ K\end{pmatrix} = X\Omega + X_\perp K$$

接ベクトル空間の条件を満たす Ω と K の条件を考えると、

と表せる。これを $\dot{X}^TX + X^T\dot{X} = 0$ に代入して

$$(X\Omega + X_\perp K)^TX + X^T(X\Omega + X_\perp K) = \Omega^T + \Omega = 0$$

Ω は歪対称行列の必要
K は任意の行列

という条件を得る。次元について確認すると Ω は$p\times p$歪対称行列なので独立な成分の数は $p(p-1)/2$、 Kは任意の $(n-p)\times p$ 行列なので、ベクトル空間としての次元は

$$p(n-p) + p(p-1)/2 = np - p(p+1)/2$$

となり、確かに $\mathrm{St}(p,n)$ の次元と一致する。

シュティーフェル多様体 Σ の ある点 $X^{n\times p}$ における 接ベクトル空間 の元 U は
任意の 歪対称行列 $\Omega^{p\times p}$ と 任意の行列 $K^{(n-p)\times p}$ を用いて、
$U = X\Omega + X_\perp K$ と書ける。

$$T_X\mathrm{St}(p,n) = \{X\Omega + X_\perp K \mid \Omega^T = -\Omega, K \in \mathbb{R}^{(n-p)\times p}\}$$

接ベクトル空間への直交射影の計算

$$T_X\mathrm{St}(p,n) = \{X\Omega + X_\perp K \mid \Omega^T = -\Omega, K \in \mathbb{R}^{(n-p)\times p}\}$$

であったので $St(p,n)$ の X での接ベクトルは $U = X\Omega + X_\perp K$ と書ける。これと $T_X\mathbb{R}^{n\times p} \simeq \mathbb{R}^{n\times p}$ のベクトル V の内積は

$$\langle U, V\rangle = \mathrm{tr}(U^T V) = \mathrm{tr}(\Omega^T X^T \overset{V}{\cancel{Y}}) + \mathrm{tr}(K^T X_\perp^T \overset{V}{\cancel{Y}})$$

接ベクトルUに直交するベクトルVを求めることで、直交補空間が求まる.

となる。これが任意の U に対して 0 になる為にはまず $X_\perp^T \overset{V}{\cancel{Y}} = 0$ となる事が必要で、その為には $\overset{V}{\cancel{Y}} = XS$ と表せる事が必要十分条件である(
$\because \mathrm{span}(\overset{V}{\cancel{Y}}) \subset \mathrm{span}(X_\perp)^\perp = \mathrm{span}(X)$であれば良いから)

↖ $V=XS$ であれば $\langle U,V\rangle$ の第2項は 0

$V=XS$, $X^TX=I$ を利用

すると $\langle U, V\rangle = \mathrm{tr}(\Omega^T S)$ となるが Ω^T は歪対称行列であるのでこれが、これが任意の Ω に対して 0 となる必要十分条件は S が対称行列である事。 よって $T_X\mathrm{St}(p,n)$ の法線ベクトル空間は

$\langle U,V\rangle$ の第2項を消すために $V=XS$

$$(T_X\mathrm{St}(p,n))^\perp = \{\underbrace{XS}_{=V} \mid S \in \mathrm{Sym}_p\}$$

$\langle U,V\rangle$ の第1項を消すために S は対称行列

$\langle U,V\rangle$: 接ベクトル, 直交補空間

となる。$V \in \mathbb{R}^{n\times p}$ の直交分解は、

$$\mathrm{sym}(A) = \frac{A + A^T}{2}, \mathrm{skew}(A) = \frac{A - A^T}{2}$$

とおくと $Y \in \mathbb{R}^{n\times p}$ に対して 接ベクトル空間への直交射影 $P_X(Y)$, 接ベクトル空間の直交補空間への射影 $P_X^\perp(Y)$

$$P_X(\overset{Y}{\cancel{V}}) = (I - XX^T)\overset{Y}{\cancel{V}} + X\mathrm{skew}(X^T \overset{Y}{\cancel{V}})$$
$$P_X^\perp(\overset{Y}{\cancel{V}}) = X\mathrm{sym}(X^T \overset{Y}{\cancel{V}})$$

この形になることは逆算的にしか分からない?

となる。実際に計算してみると $\mathrm{sym}(A) + \mathrm{skew}(A) = A$ であるから

$$P_X(\overset{Y}{\cancel{V}}) + P_X^\perp(\overset{Y}{\cancel{V}}) = (I - XX^T)\overset{Y}{\cancel{V}} + XX^T\overset{Y}{\cancel{V}} = \overset{Y}{\cancel{V}}$$

であり、内積を計算すると

$$\langle P_X^\perp(\overset{Y}{\cancel{V}})^T, P_X(\overset{Y}{\cancel{V}})\rangle = \mathrm{tr}(\mathrm{sym}(X^T\overset{Y}{\cancel{V}})\mathrm{skew}(X^T\overset{Y}{\cancel{V}})) = 0$$

となって確かに直交する。

◎ 勾配ベクトル $\nabla f(x)$ を 接ベクトル空間 $T_X St(p,n)$ へ射影する場合.

$$\begin{aligned}
\mathrm{grad} f &= P_X(\nabla f(x)) \\
&= \nabla f(x) - P_X^\perp(\nabla f(x)) \\
&= \nabla f(x) - X\,\mathrm{sym}(X^T \nabla f(x)) \\
&= \nabla f(x) - \frac{1}{2}\{XX^T\nabla f(x) + X\nabla f(x)^T X\}
\end{aligned}$$

例9.9: シュティーフェル多様体のQR分解によるレトラクション

シュティーフェル多様体 $\mathrm{St}(p,n)$ の点 X と $V \in T_X\mathrm{St}(p,n)$ に対して

$$R_X(V) = \mathrm{qf}(X + V)$$

$\mathrm{qf}(A)$ は行列AをQR分解したときのQ行列を返す関数

アルゴリズム 5.2 リーマン多様体 M 上の曲線上の探索を用いる最適化アルゴリズム

初期点 $x_0 \in M$ を選ぶ.
for $k = 0, 1, 2, \ldots$ **do**
  探索方向 $\eta_k \in T_{x_k}M$ とステップ幅 $t_k > 0$ を求める.
  次の点 x_{k+1} を $x_{k+1} := R_{x_k}(t_k\eta_k)$ によって計算する.
end for

シュティーフェル多様体上の最適化

$$\max_{X} \ f(X)$$
$$\text{s.t.} \quad X^T X = I$$

while (収束するまで)

$$\nabla f(X)$$

$\downarrow$ 接ベクトル空間へ射影

$$\mathrm{grad} f = \nabla f(X) - \frac{1}{2}\left\{ X X^T \nabla f(X) + X \nabla f(X)^T X \right\}$$

$\downarrow$ 解の更新 (レトラクション)

$$X_{k+1} = R_{X_k}\left(t_k \, \mathrm{grad} f\right)$$
$$= \mathrm{qf}\left(X_k + t_k \, \mathrm{grad} f\right)$$

min の場合 マイナス

max の場合 プラス

レイリー商 (シュティーフェル多様体上の最適化)

$$\max_{X \in \mathbb{R}^{m\times n}} \quad f(X) = \mathrm{tr}(X^T A X)$$

$$\text{s.t.} \quad X^T X = I_n \quad (n \leq m)$$

Aは半正定値行列

$(A \in S^{m\times m})$

$$\mathrm{tr}(X^T A X) = \sum_{i=1}^{n} x_i^T A x_i$$

$$X = [x_1 \cdots x_n]$$

x_iは最大固有値に対応する固有ベクトルを並べれば最適

$$\nabla f(X) = \frac{d}{dX}\mathrm{tr}(X^T A X) = (A + A^T) X$$

while (収束するまで)

$\nabla f(X)$

↓ 接ベクトル空間へ射影

$$\mathrm{grad} f = \nabla f(X) - \frac{1}{2}\{X X^T \nabla f(X) + X \nabla f(X)^T X\}$$

↓ 解の更新 (レトラクション)

$$X_{k+1} = R_{X_k}(t_k \, \mathrm{grad} f)$$

$$= \mathrm{qf}(X_k + t_k \, \mathrm{grad} f)$$

最適解は $X = U$ $(A = U\Lambda U^T)$

最大値は $f(X) = \sum_{n=1}^{m} \lambda_n$

複素数の場合

X^*の微分が必要

$$\nabla f = \frac{d}{dX^*}\mathrm{tr}(X^H A X) = \frac{d}{dX^*}\mathrm{tr}((AX)^T X^*) = AX$$

$$\mathrm{grad} f = \nabla f(X) - \frac{1}{2}\{X X^H \nabla f(X) + X \nabla f(X)^H X\}$$

$$X_{k+1} = \mathrm{qf}(X_k + t_k \, \mathrm{grad} f)$$

複素数の場合 のスカラー値関数の行列微分

$$f: \underset{z}{\mathbb{C}^{M\times N}} \to \underset{f(z)}{\mathbb{C}}$$

$$df = \mathrm{tr}\left\{\left(\frac{\partial f}{\partial z}\right)^T dz\right\} + \mathrm{tr}\left\{\left(\frac{\partial f}{\partial z^*}\right)^T dz^*\right\}$$

zとz*を独立な変数とみなせる

$$z^c = \begin{bmatrix} z \\ z^* \end{bmatrix}$$

$$f(z^c + \Delta z^c) = f(z^c) + \left(\frac{\partial f}{\partial z^c}\right)^T \Delta z^c + \cdots$$

$$df = \left(\frac{\partial f}{\partial z^c}\right)^T dz^c$$

$$= \left(\frac{\partial f}{\partial z}\right)^T dz + \left(\frac{\partial f}{\partial z^*}\right)^T dz^*$$

複素数での勾配降下法について

$\frac{\partial f}{\partial \bar{z}} = \overline{\left(\frac{\partial f}{\partial z}\right)}$ のとき

$$f(z^c + \Delta z^c) = f(z^c) + 2\mathrm{Re}\left\{\left(\frac{\partial f}{\partial z}\right)^T \Delta z\right\}$$

- 複素領域でのテイラー展開 (1次)

$$f(z^c + \Delta z^c) = f(z^c) + \left(\frac{\partial f(z^c)}{\partial z^c}\right)^T \Delta z^c$$

$$= f(z^c) + \left[\left(\frac{\partial f}{\partial z}\right)^T \left(\frac{\partial f}{\partial \bar{z}}\right)^T\right] \begin{bmatrix} \Delta z \\ \Delta \bar{z} \end{bmatrix}$$

$$= f(z^c) + \left(\frac{\partial f}{\partial z}\right)^T \Delta z + \left(\frac{\partial f}{\partial \bar{z}}\right)^T \Delta \bar{z}$$

$f(z)$を最大化するΔzの方向は $\Delta z = \left(\frac{\partial f}{\partial z}\right)^*$

勾配降下(上昇)でとるΔzを動かす理由,

$\frac{\partial f}{\partial z}$をそのまま使うのではなく、共役をとる必要がある

勾配降下のステップ

$$z_{k+1} = z_k - \alpha\left(\frac{\partial f}{\partial z}(z_k)\right)^*$$