

双対平坦・双直交 Dually flat, dually orthogonal

ryamada

2017年3月5日

- 1 測地線 Geodesic
- 2 平坦 Flatness
 - 2.1 平坦とフィッシャー情報量と接続係数 Flatness, Fisher information and connection coefficient
- 3 二重平坦 Dually flat
 - 3.1 正規分布の例 Example; Normal distribution
- 4 双対座標系 Dual coordinate system
 - 4.1 正規分布の例 Example, normal distribution

1 測地線 Geodesic

- 飛行機で遠くへ、たとえば大阪からロサンゼルスへ、行くとき、両都市の緯度は大体同じなので、その緯線に沿って飛ぶかと言えば、そんなことはなく、北東に向かって飛び始めて、だんだん真東へと向き、南東に向きを変えて到着する。Imagine a flight from Osaka to Los Angeles, whose latitudes are similar. The course is not along the latitude line but it heads north-east first then heads to east and changes the direction to south-east.
- それが緯線に沿うよりも短距離だから。The flight course is the shortest path.
- 両都市を通る大円に近いコースを取っている。The flight path is along the great circle on which two cities are.
- この大円コースが球面での「まっすぐな線」である。This great circle course is "the straight line" on the sphere.
- したがって、地球表面に緯度・経度という座標の取り方(パラメタの取り方)をしたときには、「まっすぐな線」というのは、座標系に対しては曲線となることを意味している。From this, we can tell that straight lines are curved in the coordinate system of latitude and longitude.
- このような曲がった空間でのまっすぐな線が測地線。The straight lines in the curved space are geodesics.

2 平坦 Flatness

- 空間には色々な座標の取り方があるが、ある座標系を取って、その座標をきれいな格子と考えて、地図を描き、その地図において、いわゆる直線で進むことにする。前の例では、緯線・経線を引いて、緯線に沿って進む、というような進み方である。We can assign various coordinate systems to a space. Assume you take paths that are linearly expressed in a coordinate system.
- この進み方が、「たまたま」測地線という意味でも「まっすぐな線」であるとき、この座標系の取り方は、「平坦」である、と言う。If the paths appearing straight in the coordinate system are geodesics, then the coordinate system is flat.

2.1 平坦とフィッシャー情報量と接続係数

Flatness, Fisher information and connection coefficient

- フィッシャー情報量は、局所の長さ・内積の測り方を決めている。Fisher information defines local elongation and inner product.
- フィッシャー情報量は座標系の取り方に依じて表現される。Fisher information's expression depends on the coordinate systems.
- フィッシャー情報量は局所の長さを決めているので、曲線を引くときにその長さをどのくらいと見積もるかを定める。Fisher information tells local length, therefore it determines the length of curves/lines in the space because the length is integral of local lengths.
- したがって、フィッシャー情報量が(フィッシャー情報量を構成する関数の変化具合が)曲線に沿ってどう変化するかを検討することで、どのような座標系の取り方は、平坦なのかが解る。This means that the changes of Fisher information along a curve have information whether the curves are geodesics and also have information on flatness of coordinate system.
- 実際、平坦ならば、0になるべき要素を接続係数 $F_{ki,j}^{(\alpha)}$ と呼ぶが、それは以下で示されることからその様子が見て取れる。Actually connection coefficient $F_{ki,j}^{(\alpha)}$ is defined so that $F_{ki,j}^{(\alpha)}$ should be zero when flat. The formula below implicates that connection coefficients are closely related to Fisher information.

$$\begin{aligned}
 g(ij) &= \int \frac{\partial 2\sqrt{p}}{\partial \theta_i} \frac{\partial 2\sqrt{p}}{\partial \theta_j} dx \\
 &= \int \frac{\partial l^{(\alpha)}}{\partial \theta_i} \frac{\partial l^{(-\alpha)}}{\partial \theta_j} dx \\
 F_{ki,j}^{(\alpha)} &= \int \frac{\partial}{\partial \theta_k} \frac{\partial l^{(\alpha)}}{\partial \theta_i} \frac{\partial l^{(-\alpha)}}{\partial \theta_j} dx
 \end{aligned}$$

- ただし、 $l^{(\alpha)}$ はフィッシャー情報量をうまく2つの関数の偏微分の積を使って表すために登場した関数。where $l^{(\alpha)}$ stands for function that expresses Fisher information as product of partial derivatives of them.

$$\begin{aligned}
 l^{(\alpha)} &= \frac{2}{1-\alpha} p^{\frac{1-\alpha}{2}}; \alpha \neq 1 \\
 &= \log p; \alpha = 1
 \end{aligned}$$

- 接続係数が0になるのは、 $\alpha = \pm 1$ のときであることが示せる。そして $\alpha = \pm 1$ の2つはそれぞれ Connection coefficients are 0 only when $\alpha = \pm 1$.
- そして、確率分布の多くが、このような座標系のペアを持てることが知られているので、そのような確率分布を配した空間・多様体は二重平坦である、と言われる。Almost all probability distributions are known to have a pair of coordinate systems that are both flat in the space of distributions. And the manifolds are called dually flat.

3 二重平坦 Dually flat

3.1 正規分布の例 Example; Normal distribution

- 平均 m 、標準偏差 s で表現すれば、Formula of normal distribution with mean m and standard deviation s .

$$p(x|m, s) = \frac{1}{\sqrt{2\pi}s} e^{-\frac{(x-m)^2}{2s^2}}$$

- これに (m, s) という二次元座標を与えた、と見るができる。This formula can be read that 2d coordinates (m, s) are given to the distribution.
- 今、異なる2つの座標系 $(\theta_1, \theta_2), (\eta_1, \eta_2)$ を与える。We take two different coordinate systems $(\theta_1, \theta_2), (\eta_1, \eta_2)$ as below.
- それぞれ、 $\alpha = 1, \alpha = -1$ に対応する2つの座標系として知られているものである。They are known as coordinate systems corresponding to $\alpha = 1, \alpha = -1$.

$$\theta_1 = \frac{m}{s^2}$$

$$\theta_2 = -\frac{1}{2s^2}$$

$$m = -\frac{\theta_1}{2\theta_2}$$

$$s^2 = -\frac{1}{2\theta_2}$$

$$\eta_1 = m$$

$$\eta_2 = m^2 + s^2$$

$$m = \eta_1$$

$$s^2 = \eta_2 - \eta_1^2$$

- (θ_1, θ_2) の格子を描く Let's draw lattice of (θ_1, θ_2)
- 位置によって、格子が作る伸び縮みの具合は変化しているし、直交しているわけでもない The lattice is deformed and the features of elongation varies among locations and the angles are not perpendicular.
- 今、描図に用いている (m, s) が絶対的な座標系ではないので、伸び縮みの変化や角度が「本当のところ」はどうなっているのかは、図からは分からない。それはフィッシャー情報量を各所で調べることによってわかる The coordinate system (m, s) is one of many coordinate systems and it is not the right one. Therefore this drawing does not tell "real" elongation features of (θ_1, θ_2) system. If you want to know the "real" elongation, you have to check its Fisher information.

```

# 正規分布、(m, s)座標に(theta1, theta2) 格子
# (theta1, theta2) lattice on the map with (m, s) coordiante system
# s^2 = m/theta1
# s^2 = -1/(2*theta2)
theta1 <- seq(from=1/2, to=1000, by=1/2)
theta2 <- -seq(from=1/2, to=1000, by=1/2)
fr <- matrix(c(-1, 0, 1, 1), byrow=TRUE, 2, 2)
plot(fr, asp=1, col=0, ylim=c(0, 2), xlab="m", ylab = "sd", main="theta lattice")

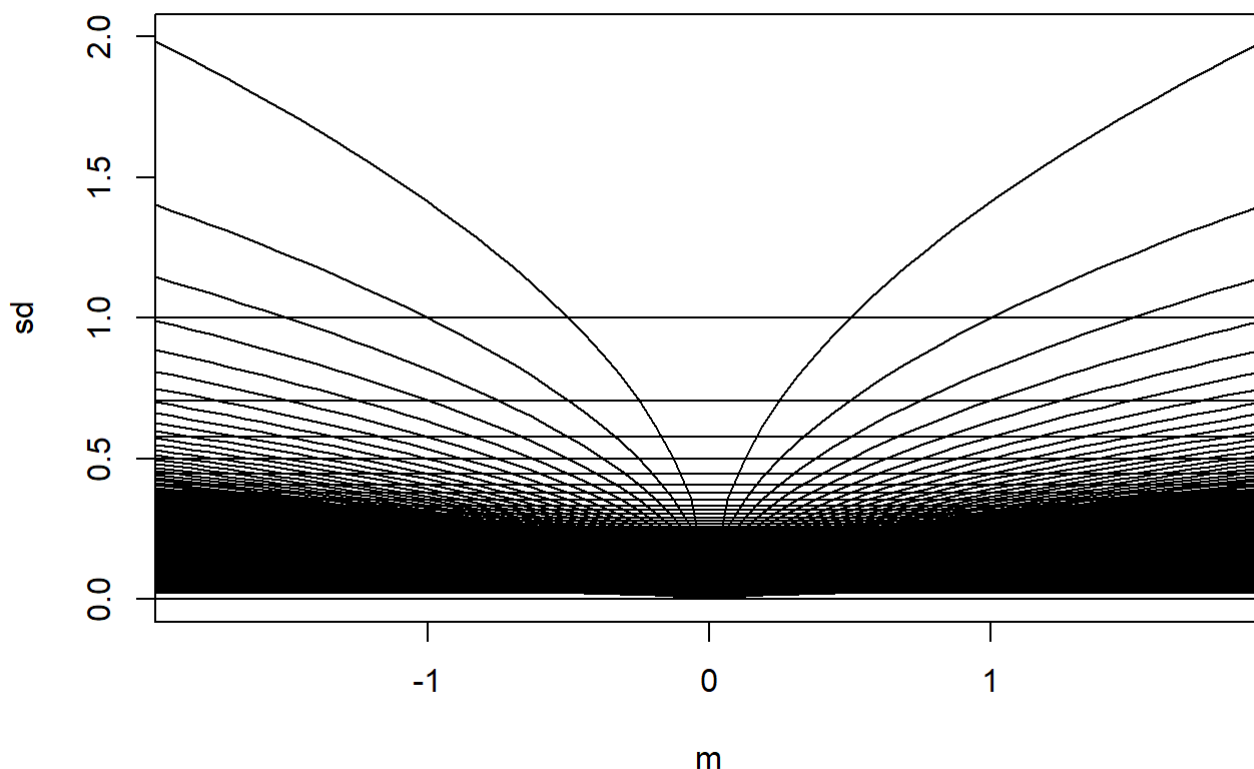
t <- seq(from=0, to=1, length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t, sqrt(t/theta1[i]), type="l")
  points(-t, sqrt(t/theta1[i]), type="l")
}

for(i in 1:length(theta2)){
  abline(h=sqrt(-1/(2*theta2[i])))
}
abline(h=0)

```

theta lattice



- (η_1, η_2) の格子を描く Let's draw lattice of (η_1, η_2)
- 位置によって、格子が作る伸び縮みの具合は変化しているし、直交しているわけでもない Elongation feaures depends on location. Not perpendicular.
- 今、描図に用いている (m, s) が絶対的な座標系ではないので、伸び縮みの変化や角度が「本当のところ」はどうなっているのかは、図からは分からない。それはフィッシャー情報量を各所で調べることによってわかる For the true elongation, check Fisher information.

```

# 正規分布、(m, s)座標に(eta1, eta2)格子
# (eta1, eta2) lattice on the map with (m, s) coordiante system

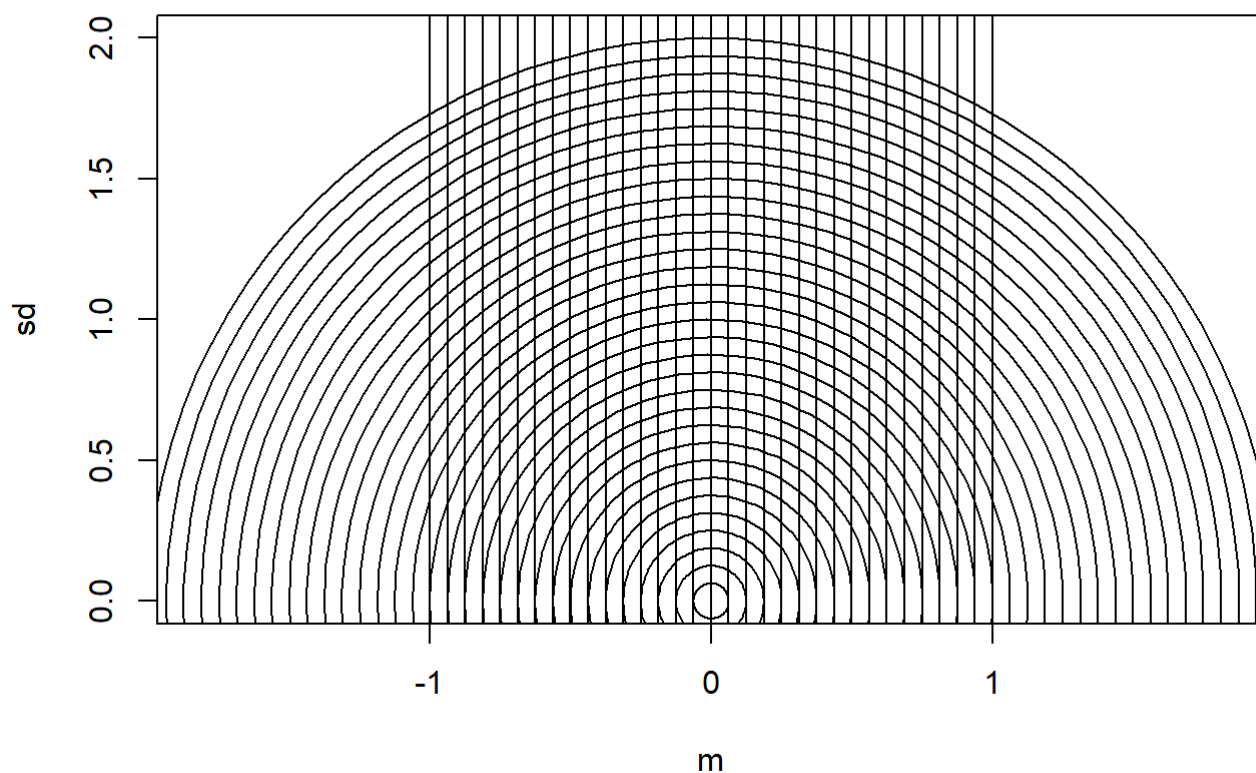
# m = eta1
# m^2+s^2=eta2
eta1 <- seq(from=-1, to=1, by=1/(2^4))
eta2 <- seq(from=0, to=2, by=1/(2^4))

fr <- matrix(c(-1, 0, 1, 1), byrow=TRUE, 2, 2)
plot(fr, asp=1, col=0, ylim=c(0, 2), xlab="m", ylab = "sd", main="eta lattice")

for(i in 1:length(eta1)){
  abline(v=eta1[i])
}
t <- seq(from=0, to=1, length=100) * 2*pi
for(i in 1:length(eta2)){
  points(eta2[i] * cos(t), eta2[i] * sin(t), type="l")
}

```

eta lattice



- (θ_1, θ_2) と (η_1, η_2) の2種類の格子を重ねてみる Draw two coordinate system lattice together.

```

fr <- matrix(c(-0.5, 0, 0.5, 0.5), byrow=TRUE, 2, 2)
plot(fr, asp=1, col=0, ylim=c(0, 1), xlab="m", ylab = "sd", main="theta & eta lattice")
t <- seq(from=0, to=1, length=100) * 2*pi

for(i in 1:length(theta1)){
  points(t, sqrt(t/theta1[i]), type="l", col=2)
  points(-t, sqrt(t/theta1[i]), type="l", col=2)
}

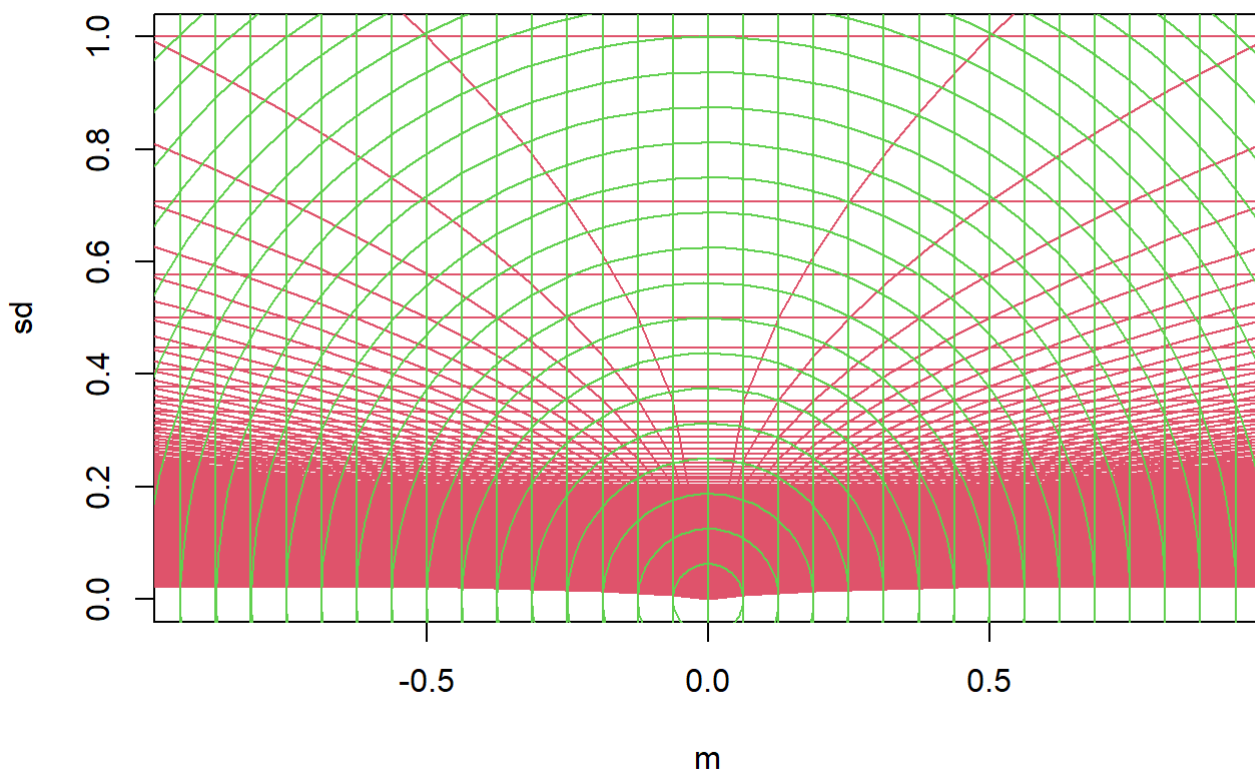
for(i in 1:length(theta2)){
  abline(h=sqrt(-1/(2*theta2[i])), col=2)
}

for(i in 1:length(eta1)){
  abline(v=eta1[i], col=3)
}

t <- seq(from=0, to=1, length=100) * 2*pi
for(i in 1:length(eta2)){
  points(eta2[i] * cos(t), eta2[i] * sin(t), type="l", col=3)
}

```

theta & eta lattice



- θ_1 と η_1 とを同じ色で、 θ_2 と η_2 とを同じ色で描きなおしてみる Draw the same but colors are common for θ_i and η_i not for θ_1 と θ_2 .

```

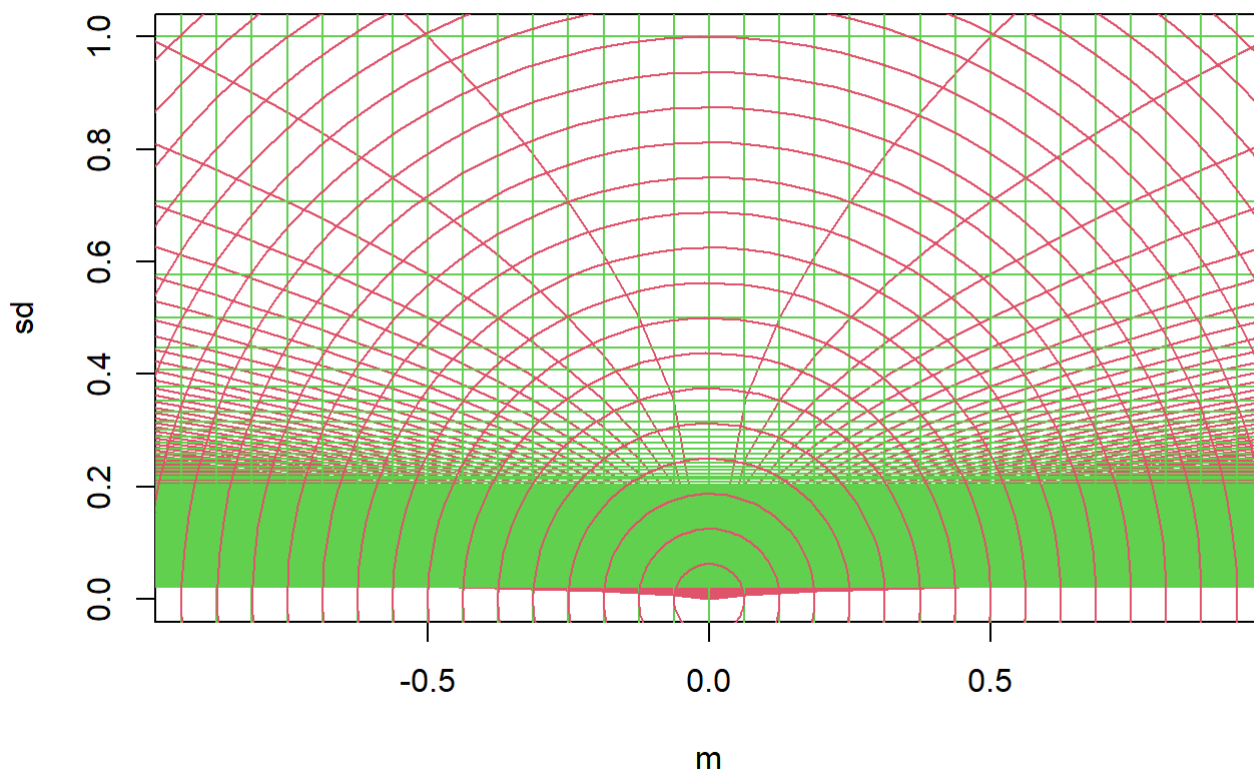
fr <- matrix(c(-0.5, 0, 0.5, 0.5), byrow=TRUE, 2, 2)
plot(fr, asp=1, col=0, ylim=c(0, 1), xlab="m", ylab = "sd", main="theta & eta lattice")
t <- seq(from=0, to=1, length=100) * 2*pi

for(i in 1:length(theta1)) {
  points(t, sqrt(t/theta1[i]), type="l", col=2)
  points(-t, sqrt(t/theta1[i]), type="l", col=2)
  #segments(0, 0, 2, 2/theta1[i], col=2)
  #segments(0, 0, -2, 2/theta1[i], col=2)
}

for(i in 1:length(theta2)) {
  abline(h=sqrt(-1/(2*theta2[i])), col=3)
}
for(i in 1:length(eta1)) {
  abline(v=eta1[i], col=3)
}
t <- seq(from=0, to=1, length=100) * 2*pi
for(i in 1:length(eta2)) {
  points(eta2[i] * cos(t), eta2[i] * sin(t), type="l", col=2)
}

```

theta & eta lattice



4 双対座標系 Dual coordinate system

- 正規分布は2次元多様体であるので、2次元座標を与えられる Coordinate systems of normal distribution is 2d.
- 二つの座標系 (θ_1, θ_2) , (η_1, η_2) を示した Two systems (θ_1, θ_2) and (η_1, η_2) were introduced above.

- この座標系は、正規分布を持つ特別な座標系ペアである They are special for normal distribution.
- どちらも平坦 Both are flat.
- 相互に直交(双直交) They are mutually orthogonal (dually orthogonal).
- どちらも、平坦であるというのは、座標系を「信じて」まっすぐに進むとそれが測地線になること。このことは、それぞれの座標系が持つ性質であって、2つの座標系の関係を考える必要がない Both are flat, which means that straight paths in either coordinate system are geodesics. This feature is intrinsic for both system and no interaction between two systems.
- 相互に直交(双直交)であることを考えるには、それぞれの座標系の個々の要素に対応付けをする必要がある。 θ_1 と η_1 とは、ある理由で対応があり、その同じ理由で θ_2 と η_2 とも対応がある、というような関係である。一般に、 k 次元の場合も θ_i と $\eta_i, i = 1, 2, \dots, k$ に対応がある The dual orthogonality is the feature between two systems. Therefore we need a piece of rule between the two systems. The elements of θ system, and the elements of η system are in the one-to-one correspondence relation. θ_1 is something specific for η_1 but not for η_2 .
- このとき、 θ_i と η_j とに関する、偏微分ベクトル(局所の座標方向のベクトルであって、伸び縮みを表すベクトル)同士が、正規直行基底であるかのような性質を持ち、そのことを双直交と言う Now we have mutual relation among θ_i and η_j depending on $i = j$ or $i \neq j$ and the set of partial derivative vectors of θ system and the set of partial derivative vectors of η system are in the relation where the inner products of both system vectors orthonormal. And this relation is called dual orthonormal.
- 正規直行基底であるかのように、とは、 $i = j$ のときには、内積が1であり、 $i \neq j$ のときには内積が0となることである It means that the inner product of the partial derivative vectors of θ_i and η_i is 1 and the inner product of vectors of θ_i and η_j , where $i \neq j$, is 0.
- θ_i と η_j とに関する、偏微分ベクトルと書いたが、 θ_i のそれは $l^{(\alpha=1)} = \log p$ から来る座標系なので、その偏微分ベクトルは $\frac{\partial \log p}{\partial \theta_i}$ である
- 他方 η_i のそれは $l^{(\alpha=-1)} = p$ から来る座標系なので、その偏微分ベクトルは $\frac{\partial p}{\partial \eta_i}$ 。The partial derivative vectors are somewhat different between θ system and η system. For θ system, differentiate $\log p$ and for η system, differentiate p itself.
- 別の言い方もできる。 $\int \frac{\partial \log p}{\partial \theta_i} \frac{\partial p}{\partial \eta_i} dx$ は $\int \frac{\partial \log p}{\partial \theta_i} \frac{\partial \log p}{\partial \eta_i} \times p dx$ と変形できるから、 $\log p$ の θ_i と η_j とのそれぞれの偏微分の期待値が1または0になる関係でもある Another explanation is also possible. $\int \frac{\partial \log p}{\partial \theta_i} \frac{\partial p}{\partial \eta_i} dx$ can be transformed to $\int \frac{\partial \log p}{\partial \theta_i} \frac{\partial \log p}{\partial \eta_i} \times p dx$. This means $\log p$ is partially differentiated with θ_i and η_j and their "expected value" is 1 or 0.

4.1 正規分布の例 Example, normal distribution

- 具体例で確認する Confirm what the above descriptions mean with an example.
- ある正規分布 $N(m, s^2)$ を考え、それを、少し変化させる。変化させる方向は $\theta_1, \theta_2, \eta_1, \eta_2$ とすれば、それぞれの微小変化によって正規分布が変わる、 $x \in R$ 全体にわたって増えたり減ったりする Take an instance of normal distribution $N(m, s^2)$ and change its parameters a bit. The direction of change is in $\theta_1, \theta_2, \eta_1$ or η_2 . With the change in any one of four directions, values of normal distribution for $x \in R$ changes everywhere.
- その変化の総和(x に関する積分)をパラメタの増分で割ったものが、 $\frac{\partial \log p}{\partial \theta_i}$ だったり、 $\frac{\partial p}{\partial \eta_j}$ となる The integral of the changes throughout x should divided by the change in parameter value, that is approximation of derivative.

- それを確かめるために、 (m, s) と $(\theta_1, \theta_2), (\eta_1, \eta_2)$ とを変換する関数や、 (θ_1, θ_2) のときの $\log N(m, s)$ の値や (η_1, η_2) のときの $N(m, s)$ の値を算出する関数を作る First make some utility functions to perform this numeric experiments.
- 座標の相互変換の関数 Functions to convert coordinates among three systems.

```
my.ms2theta <- function(m, s) {
  theta1 <- m/(s^2)
  theta2 <- -1/(2*s^2)
  return(c(theta1, theta2))
}

my.ms2eta <- function(m, s) {
  eta1 <- m
  eta2 <- s^2+m^2
  return(c(eta1, eta2))
}

my.theta2ms <- function(theta1, theta2) {
  m <- -theta1/(2*theta2)
  s <- sqrt(-1/(2*theta2))
  return(c(m, s))
}

my.eta2ms <- function(eta1, eta2) {
  m <- eta1
  s <- sqrt(eta2-eta1^2)
  return(c(m, s))
}

my.theta2eta <- function(theta1, theta2) {
  ms <- my.theta2ms(theta1, theta2)
  my.ms2eta(ms[1], ms[2])
}

my.eta2theta <- function(eta1, eta2) {
  ms <- my.eta2ms(eta1, eta2)
  my.ms2theta(ms[1], ms[2])
}
```

- それぞれの座標系で、 $\log p, p$ を返す関数を作る Functions to return $\log p$ or p for each coordinate systems.

```
# theta系はlog(p)
my.dnorm.theta <- function(x, theta1, theta2, log=TRUE) {
  ms <- my.theta2ms(theta1, theta2)
  dnorm(x, ms[1], ms[2], log=log)
}

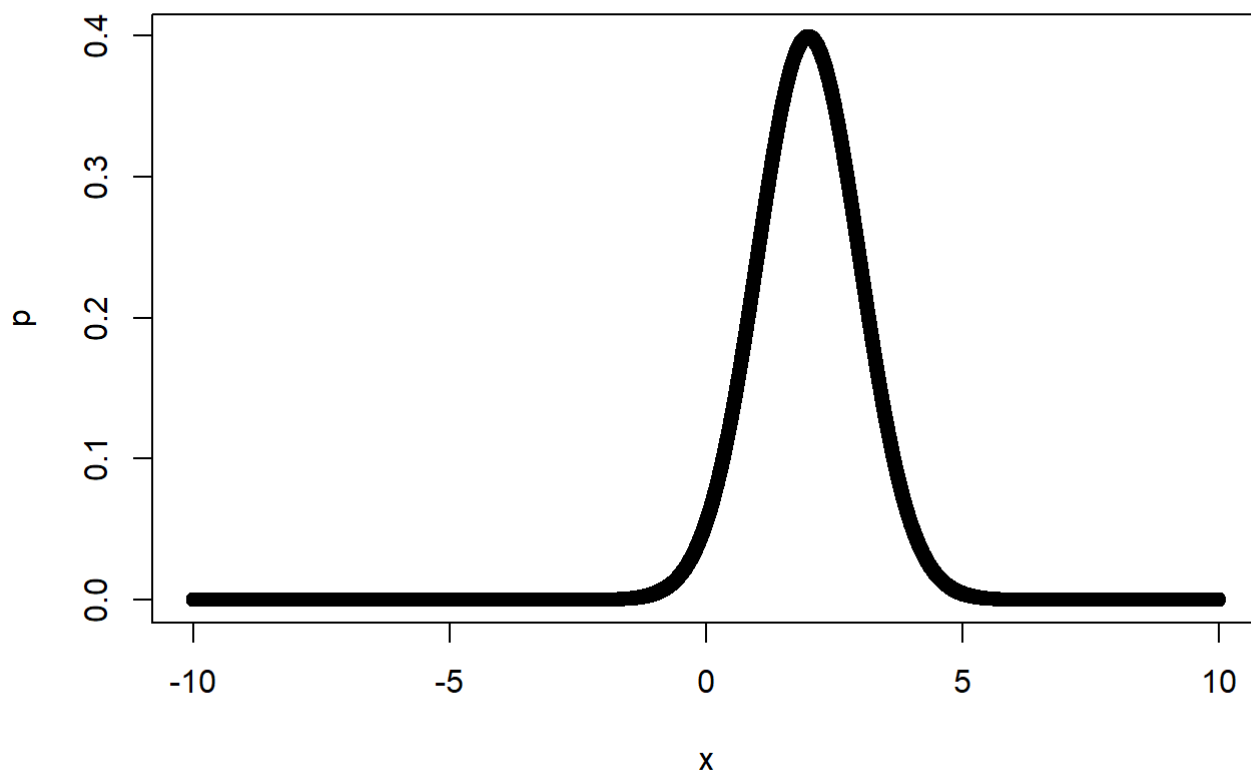
# eta系はp
my.dnorm.eta <- function(x, eta1, eta2, log=FALSE) {
  ms <- my.eta2ms(eta1, eta2)
  dnorm(x, ms[1], ms[2], log=log)
}
```

- ある特定の正規分布 $m_0 = 2, s_0 = 1$ について偏微分ベクトルを求め、双直交性を確認する Take an example normal distribution with $m_0 = 2, s_0 = 1$ and calculate numeric approximates of partial derivatives for 4 directional chnges.

```

m0 <- 2
s0 <- 1
dx <- 1/1000
x <- seq(from=-10, to=10, by=dx)
p <- dnorm(x, m0, s0)
plot(x, p)

```



- 対応する θ, η を求める θ, η ?

```

thetas <- my.ms2theta(m0, s0)
etas <- my.ms2eta(m0, s0)

```

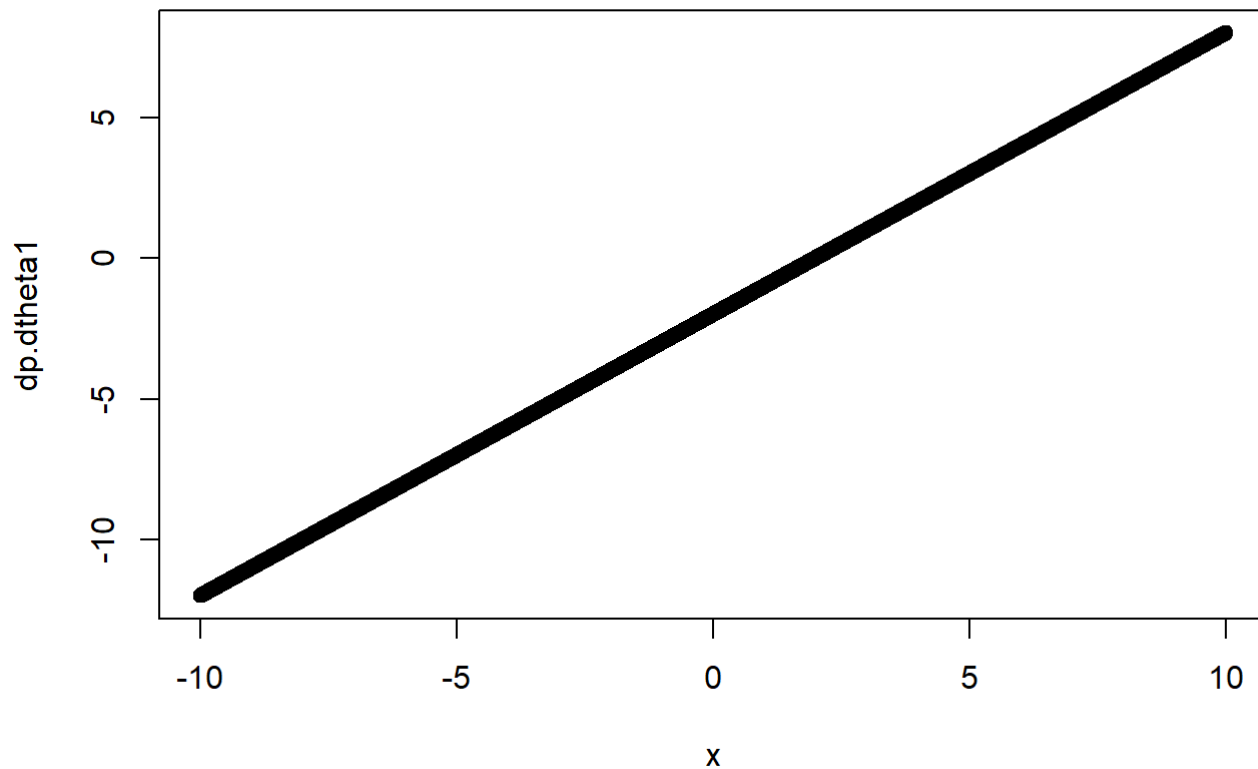
- θ_1 の偏微分を求める。その増減の様子を描く Partial derivative in the direction of θ_1 .

```

d.theta1 <- 0.0001
dp.dtheta1 <- (my.dnorm.theta(x, thetas[1]+d.theta1, thetas[2]) - my.dnorm.theta(x, thetas[1], thetas[2]))/d.theta1

plot(x, dp.dtheta1)

```



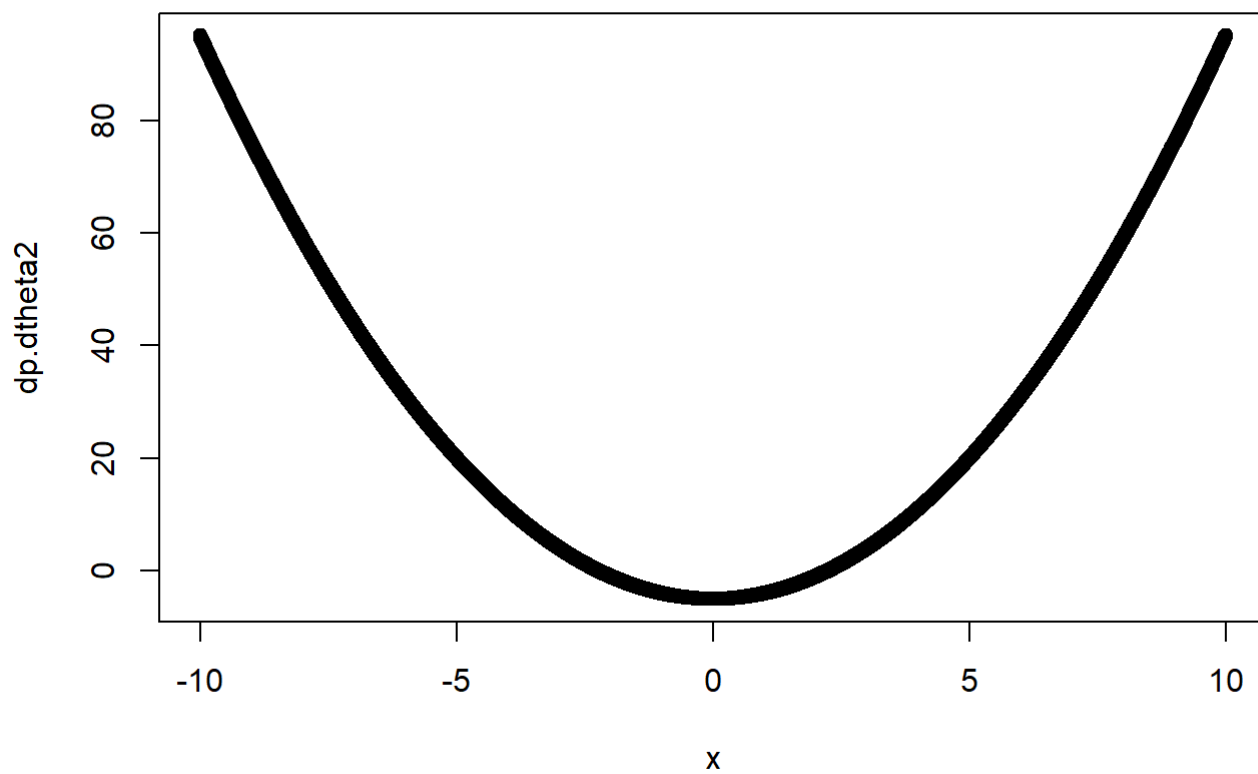
- θ_2, η_1, η_2 についても同様に求める Same for θ_2, η_1, η_2 .

```
d.theta2 <- 0.0001
```

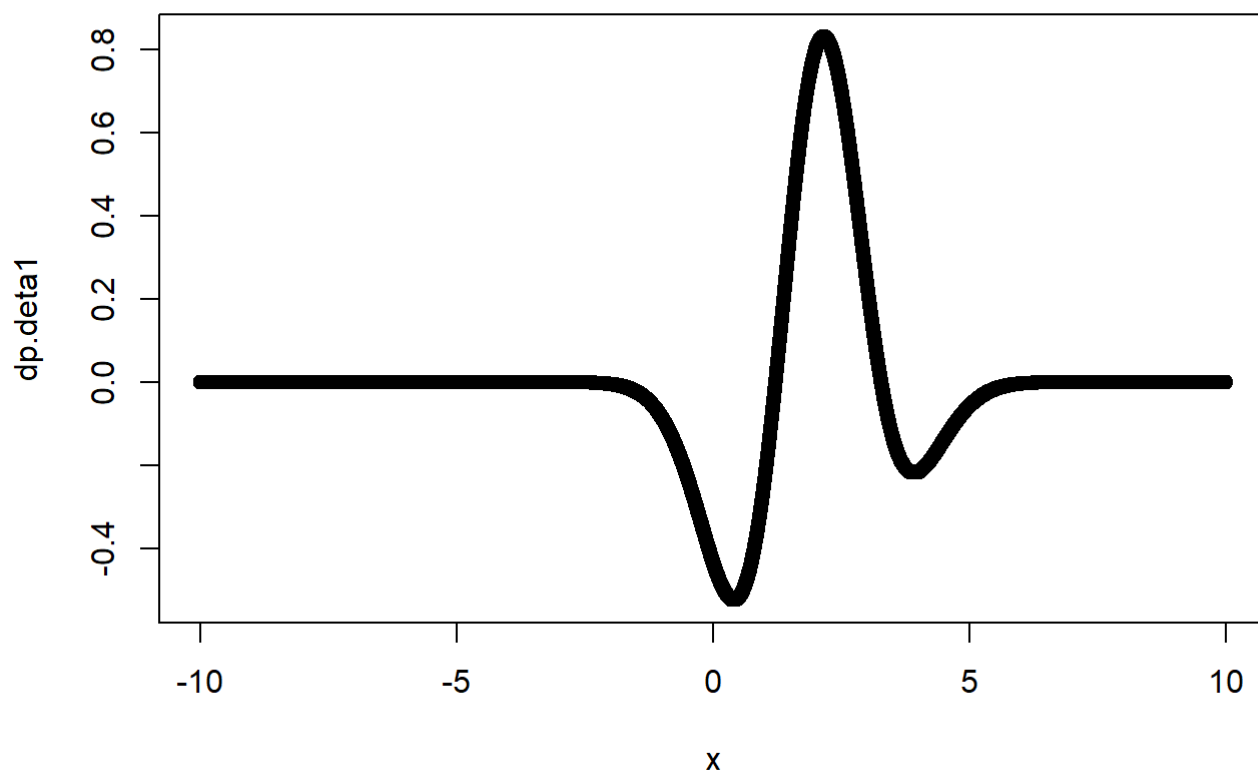
```
d.eta1 <- 0.0001
```

```
d.eta2 <- 0.0001
```

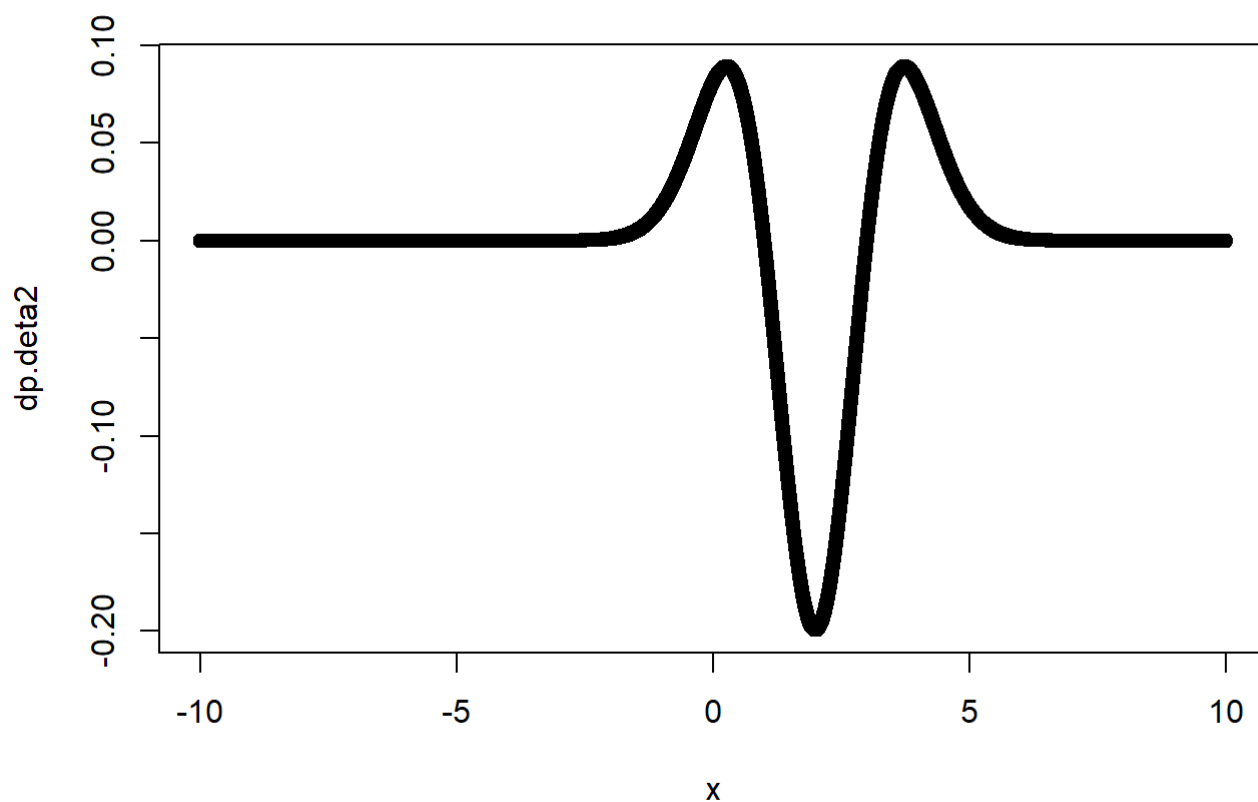
```
dp.dtheta2 <- (my.dnorm.theta(x, thetas[1], thetas[2]+d.theta2) - my.dnorm.theta(x, thetas[1], thetas[2]))/d.theta2
plot(x, dp.dtheta2)
```



```
dp.deta1 <- (my.dnorm.eta(x, etas[1]+d.eta1, etas[2]) - my.dnorm.eta(x, etas[1], etas[2]))/d.eta1  
plot(x, dp.deta1)
```



```
dp.deta2 <- (my.dnorm.eta(x, etas[1], etas[2]+d.eta2) - my.dnorm.eta(x, etas[1], etas[2]))/d.eta2
plot(x, dp.deta2)
```



- 双直交性の確認 Dual orthogonality is confirmed as;

```
sum(dp.dtheta1 * dp.deta1) * diff(x)[1]
```

```
## [1] 1
```

```
sum(dp.dtheta1 * dp.deta2) * diff(x)[1]
```

```
## [1] -1.513625e-13
```

```
sum(dp.dtheta2 * dp.deta1) * diff(x)[1]
```

```
## [1] 1.013927e-11
```

```
sum(dp.dtheta2 * dp.deta2) * diff(x)[1]
```

```
## [1] 1
```