

PRML 12.2-12.2.2

11/2/2018

酒井 一徳

- 12.2 確率的主成分分析
 - 12.2.1 最尤法による主成分分析
 - 12.2.2 EMアルゴリズムによる主成分分析

12.2 確率の主成分分析

PCA

- データから主成分を得る.
 - $(\text{データ}) = (\text{主成分}) + (\text{損失})$

PPCA

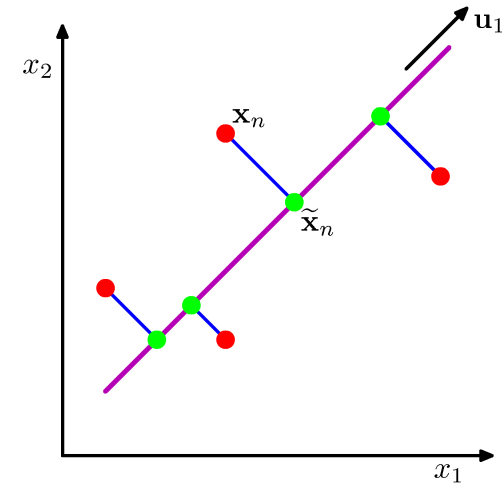
- 主成分(潜在変数)があり、データはそこにノイズが乗ったもの.
 - $(\text{主成分}) + (\text{ノイズ}) = (\text{データ})$
 - 因子分析っぽい.

PCA

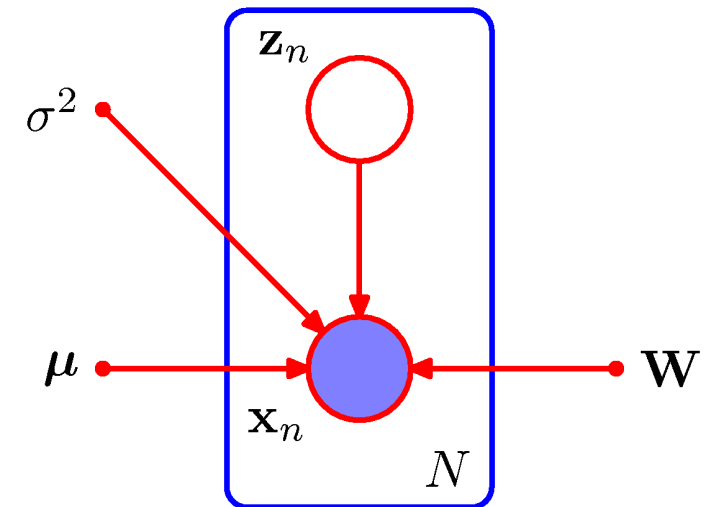
- データ空間よりも低次元の部分空間上へのデータ点の直交射影.

PPCA

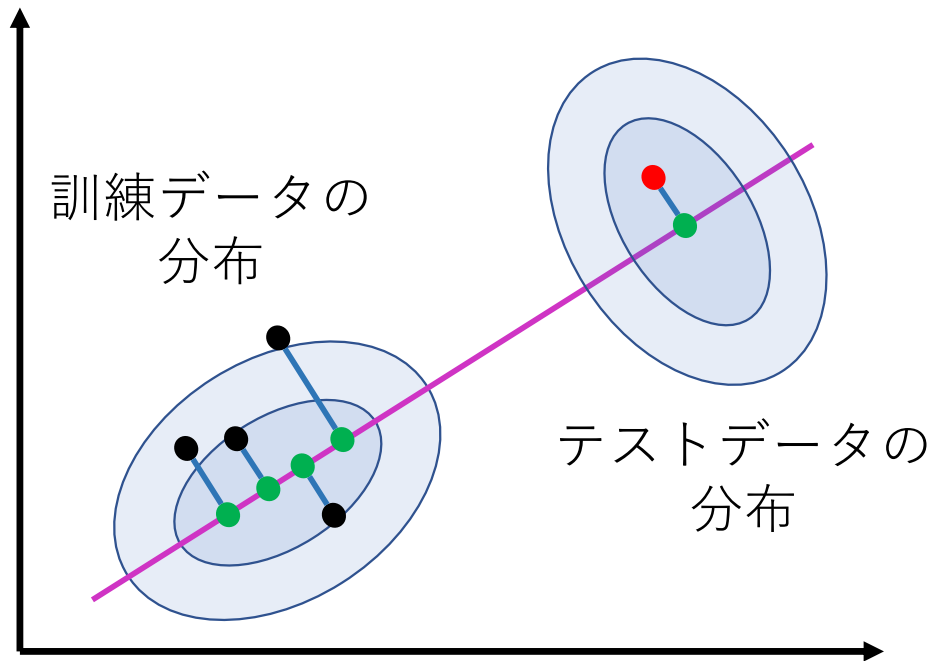
- 確率モデルは，潜在変数を持つ線形ガウスモデル.
 - 最尤解が，通常PCAを再現.



- 潜在変数モデル導入でEMの利用可.
 - 効率的.
 - 欠損値を扱える.
 - 混合モデルにも対応しやすい.
- ベイズ的取り扱いの足がかり.
 - 主部分空間の次元の自動決定(ARD).
- 生成モデルとして利用可.
- 分類問題に利用可.
- 尤度関数で他のモデルと比較が容易.



- PCAの再現コストは誤解を招く恐れがある。



$$J = \frac{1}{N} \sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|^2$$

テストデータの分布が離れていても、
小さな値になる可能性がある。

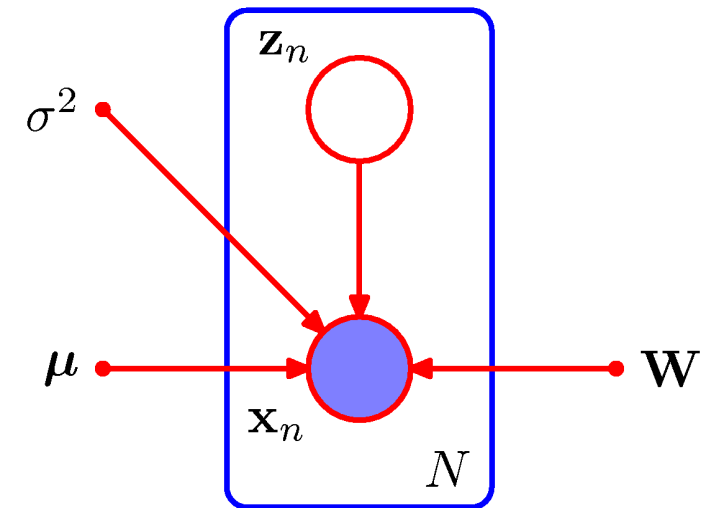
➤ 線形ガウスモデル.

➤ 潜在変数: $\mathbf{z} \in \mathbb{R}^M$
 $\mathbf{z} \sim \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$

➤ 観測変数: $\mathbf{x} \in \mathbb{R}^D$
 $\mathbf{x}|\mathbf{z} \sim \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \sigma^2\mathbf{I})$

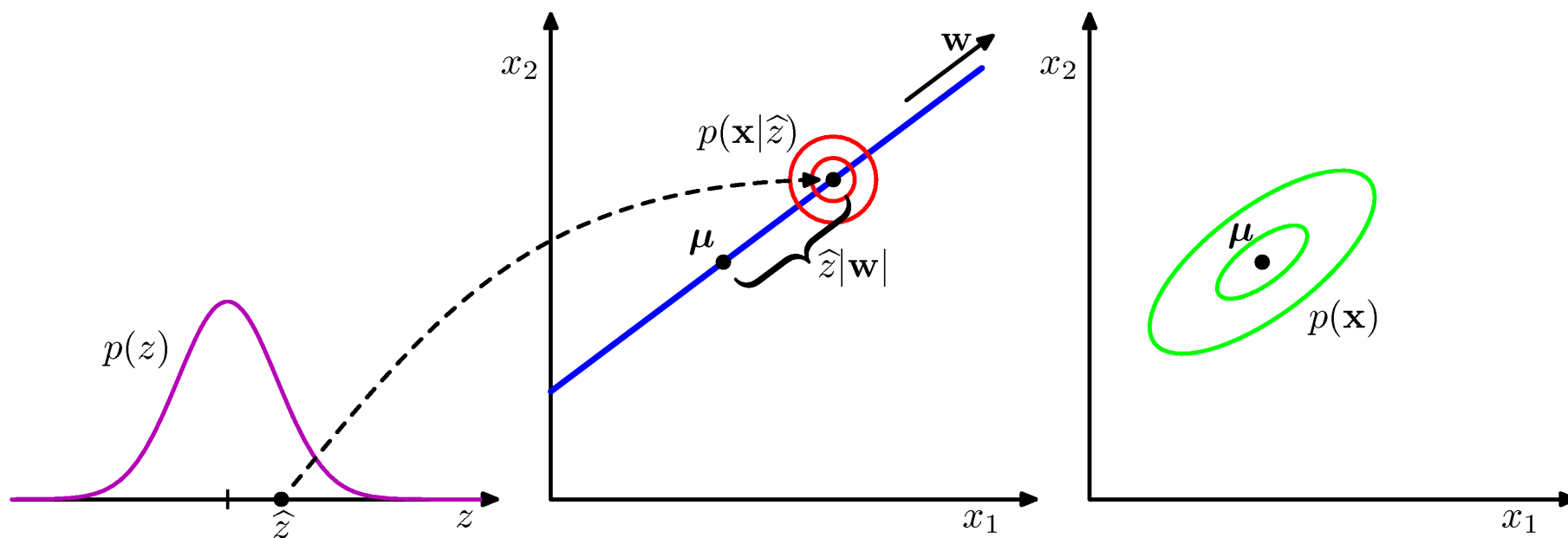
$$\mathbf{W} \in \mathbb{R}^{D \times M}$$

$$\boldsymbol{\mu} \in \mathbb{R}^D$$



$$\mathbf{x} = \mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon} \quad \text{where} \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$
$$\boldsymbol{\epsilon} \in \mathbb{R}^D$$

\mathbf{x} は \mathbf{z} の線形変換にガウス分布によるノイズを加えたものとして表現。



- \mathbf{x} の周辺分布が必要.

$$p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$$

- ガウス同士の積だから結果もガウス.

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$$

$$\mathbf{C} = \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}$$

再生性によるざっくり導出

➤ 線形変換後もガウス。

$$\begin{aligned}\tilde{\mathbf{z}} &= \mathbf{W}\mathbf{z} + \boldsymbol{\mu} \\ \mathbf{z} &\sim \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})\end{aligned}$$



$$\tilde{\mathbf{z}} \sim \mathcal{N}(\tilde{\mathbf{z}}|\boldsymbol{\mu}, \mathbf{W}\mathbf{W}^T)$$

$$\begin{aligned}\mathbf{x} &= \mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon} \\ &= \tilde{\mathbf{z}} + \boldsymbol{\epsilon}\end{aligned}$$



$$\begin{aligned}\tilde{\mathbf{z}} &\sim \mathcal{N}(\tilde{\mathbf{z}}|\boldsymbol{\mu}, \mathbf{W}\mathbf{W}^T) \\ \boldsymbol{\epsilon} &\sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})\end{aligned}$$

$$\begin{aligned}p(\mathbf{x}) &= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C}) \\ \mathbf{C} &= \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}\end{aligned}$$

$$\begin{aligned}\mathbb{E}_{p(\mathbf{x})} [\mathbf{x}] &= \mathbb{E}_{p(\mathbf{z})} [\mathbb{E}_{p(\mathbf{x}|\mathbf{z})} [\mathbf{x}]] \\ &= \mathbb{E}_{p(\mathbf{z})} [\mathbf{W}\mathbf{z} + \boldsymbol{\mu}] = \boldsymbol{\mu}\end{aligned}$$

$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \sigma^2\mathbf{I})$$

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$$

$$\begin{aligned}\text{COV}_{p(\mathbf{x})} [\mathbf{x}] &= \mathbb{E}_{p(\mathbf{z})} [\text{COV}_{p(\mathbf{x}|\mathbf{z})} [\mathbf{x}]] + \text{COV}_{p(\mathbf{z})} [\mathbb{E}_{p(\mathbf{x}|\mathbf{z})} [\mathbf{x}]] \\ &= \mathbb{E}_{p(\mathbf{z})} [\sigma^2\mathbf{I}] + \text{COV}_{p(\mathbf{z})} [\mathbf{W}\mathbf{z} + \boldsymbol{\mu}]\end{aligned}$$

$$\begin{aligned}\text{COV}_{p(\mathbf{z})} [\mathbf{W}\mathbf{z} + \boldsymbol{\mu}] &= \mathbb{E}_{p(\mathbf{z})} [(\mathbf{W}\mathbf{z} + \boldsymbol{\mu} - \mathbb{E}_{p(\mathbf{z})} [\mathbf{W}\mathbf{z} + \boldsymbol{\mu}]) (\mathbf{W}\mathbf{z} + \boldsymbol{\mu} - \mathbb{E}_{p(\mathbf{z})} [\mathbf{W}\mathbf{z} + \boldsymbol{\mu}])^T] \\ &= \mathbb{E}_{p(\mathbf{z})} [(\mathbf{W}\mathbf{z}) (\mathbf{W}\mathbf{z})^T] = \mathbf{W}\mathbf{W}^T\end{aligned}$$

$$\text{COV}_{p(\mathbf{x})} [\mathbf{x}] = \sigma^2\mathbf{I} + \mathbf{W}\mathbf{W}^T$$

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$$

$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \sigma^2\mathbf{I})$$

$$p(\boldsymbol{\epsilon}) = \mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, \sigma^2\mathbf{I})$$

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{m}, \boldsymbol{\Sigma})$$

$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \sigma^2\mathbf{I})$$

$$p(\boldsymbol{\epsilon}) = \mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, \sigma^2\mathbf{I})$$

- パラメータの再定式化で等価なモデルを得る.
- 演習12.4

$$\mathbb{E}_{p(\mathbf{x})} [\mathbf{x}] = \mathbb{E}_{p(\mathbf{z})} [\mathbf{W}\mathbf{z} + \boldsymbol{\mu}] = \mathbf{W}\mathbf{m} + \boldsymbol{\mu}$$

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{m}, \boldsymbol{\Sigma})$$

$$\text{COV}_{p(\mathbf{x})} [\mathbf{x}] = \mathbb{E}_{p(\mathbf{z})} [\sigma^2 \mathbf{I}] + \text{COV}_{p(\mathbf{z})} [\mathbf{W}\mathbf{z} + \boldsymbol{\mu}]$$



$$\begin{aligned} \text{COV}_{p(\mathbf{z})} [\mathbf{W}\mathbf{z} + \boldsymbol{\mu}] &= \mathbf{W} \mathbb{E}_{p(\mathbf{z})} [(\mathbf{z} - \mathbf{m})(\mathbf{z} - \mathbf{m})^T] \mathbf{W}^T \\ &= \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^T \end{aligned}$$

$$\text{COV}_{p(\mathbf{x})} [\mathbf{x}] = \sigma^2 \mathbf{I} + \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^T$$

➤ パラメータの再定義

$$\tilde{\boldsymbol{\mu}} = \mathbf{W}\mathbf{m} + \boldsymbol{\mu}$$

$$\tilde{\mathbf{W}} = \mathbf{W}\boldsymbol{\Sigma}^{1/2}$$



$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\tilde{\boldsymbol{\mu}}, \mathbf{C})$$

$$\mathbf{C} = \tilde{\mathbf{W}}\tilde{\mathbf{W}}^T + \sigma^2 \mathbf{I}$$

- 共分散行列 \mathbf{C} は $D \times D$ 行列.
- 逆行列 \mathbf{C}^{-1} の計算コストは $O(D^3)$.

$$\mathbf{C}^{-1} = \left(\mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I} \right)^{-1} \quad O(D^3)$$


$$\left(\mathbf{A} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C} \right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B} \left(\mathbf{D} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \right)^{-1} \mathbf{C}\mathbf{A}^{-1}$$

$$\mathbf{C}^{-1} = \sigma^{-2}\mathbf{I} - \sigma^{-2}\mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T$$

$$\mathbf{M} = \mathbf{W}^T\mathbf{W} + \sigma^2\mathbf{I}$$

$$O(M^3)$$

$$p(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z} | \mathbf{M}^{-1} \mathbf{W}^T (\mathbf{x} - \boldsymbol{\mu}), \sigma^2 \mathbf{M}^{-1})$$

$$\text{where } \mathbf{M} = \mathbf{W}^T \mathbf{W} + \sigma^2 \mathbf{I}$$

平均のみがデータに依存.

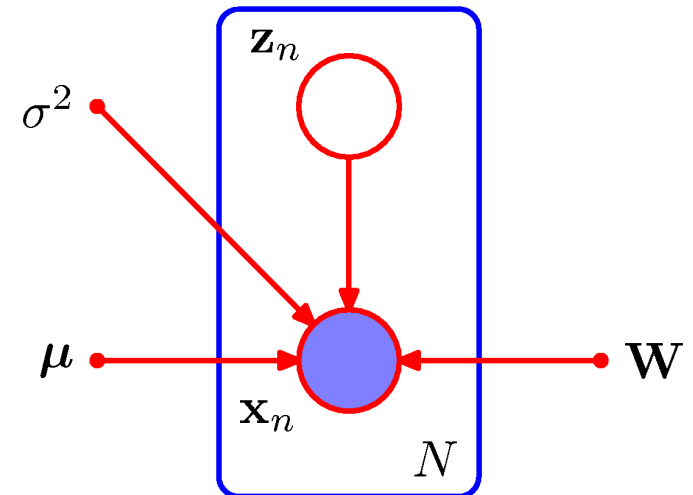
12.2.1 最尤法による主成分分析

$$\begin{aligned} \ln p(\mathbf{X}|\boldsymbol{\mu}, \mathbf{W}, \sigma^2) &= \sum_{n=1}^N \ln p(\mathbf{x}_n|\mathbf{W}, \boldsymbol{\mu}, \sigma^2) \\ &= -\frac{ND}{2} \ln(2\pi) - \frac{N}{2} \ln |\mathbf{C}| - \frac{1}{2} \sum_{n=1}^N (\mathbf{x}_n - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x}_n - \boldsymbol{\mu}) \end{aligned}$$

➤ データ集合: $\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^N$

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$$

$$\mathbf{C} = \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}$$



$$\frac{\partial}{\partial \boldsymbol{\mu}} \ln p(\mathbf{X}|\mathbf{W}, \boldsymbol{\mu}, \sigma^2) = 0 \quad \text{を解いて,}$$

$$\boldsymbol{\mu}_{\text{ML}} = \bar{\mathbf{x}} \quad \text{where} \quad \bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n$$



$$\ln p(\mathbf{X}|\mathbf{W}, \boldsymbol{\mu}, \sigma^2) = -\frac{N}{2} \left\{ D \ln(2\pi) + \ln |\mathbf{C}| + \text{Tr}(\mathbf{C}^{-1} \mathbf{S}) \right\}$$

$$\text{where} \quad \mathbf{S} = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}}) (\mathbf{x}_n - \bar{\mathbf{x}})^{\text{T}}$$

➤ 閉形式の厳密解が存在.

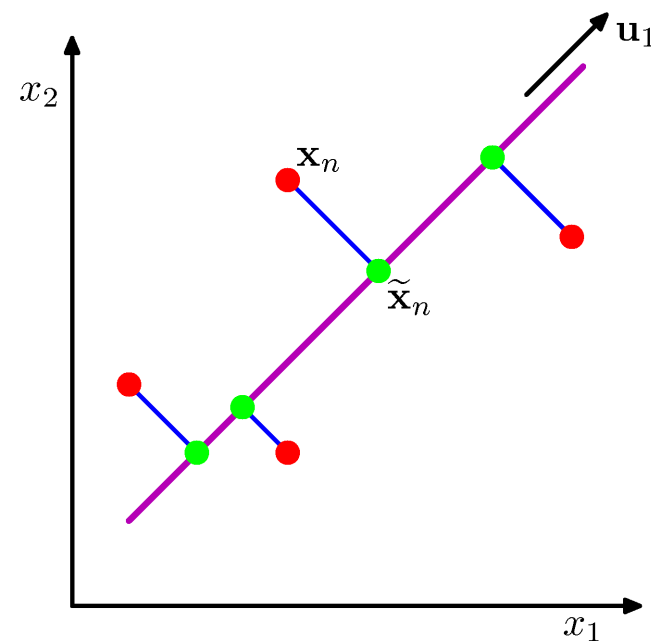
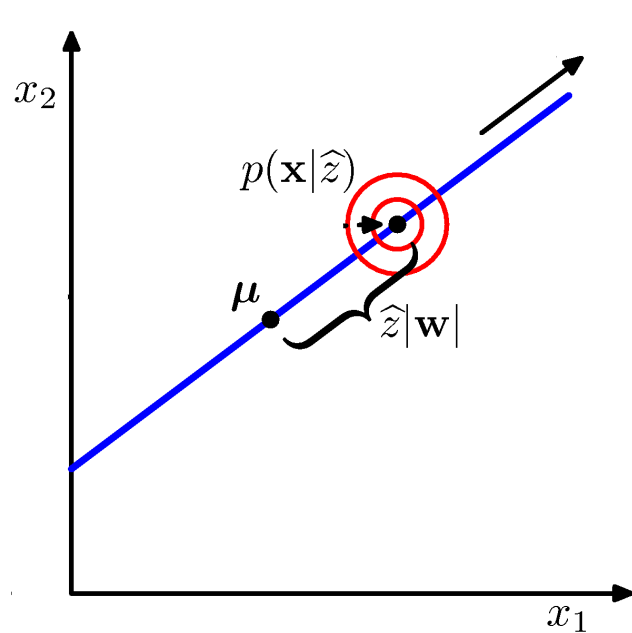
$$\mathbf{W}_{\text{ML}} = \mathbf{U}_M (\mathbf{L}_M - \sigma^2 \mathbf{I})^{1/2} \mathbf{R}$$

$$\mathbf{L}_M = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_M \end{pmatrix} \quad \begin{array}{l} : M \times M \text{ 対角行列} \\ : \text{行列 } \mathbf{S} \text{ の固有値} \\ \lambda_1 > \dots > \lambda_M \end{array}$$

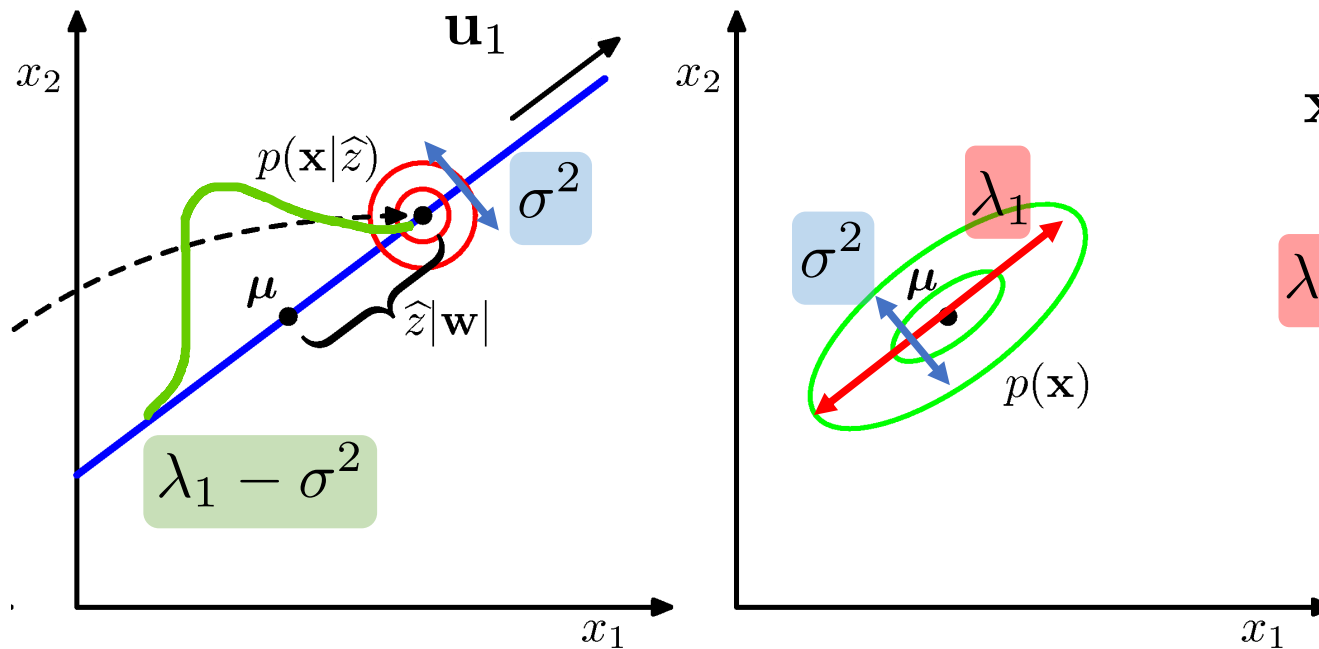
$$\mathbf{U}_M = (\mathbf{u}_1, \dots, \mathbf{u}_M) \quad \begin{array}{l} : \mathbf{u}_i \text{ は } \lambda_i \text{ に対応する固有ベクトル} \\ D \times M \text{ 直交行列} \end{array}$$

\mathbf{R} : 任意の $M \times M$ 直交行列

➤ \mathbf{W}_{ML} の列ベクトルは通常のPCAの主部分空間を成す.



- $\mathbf{R} = \mathbf{I}$ の場合: $\mathbf{W}_{\text{ML}} = \mathbf{U}_M (\mathbf{L}_M - \sigma^2 \mathbf{I})^{1/2}$
- 列ベクトルは主成分 \mathbf{u}_i を $\sqrt{\lambda_i - \sigma^2}$ で尺度変換したもの。



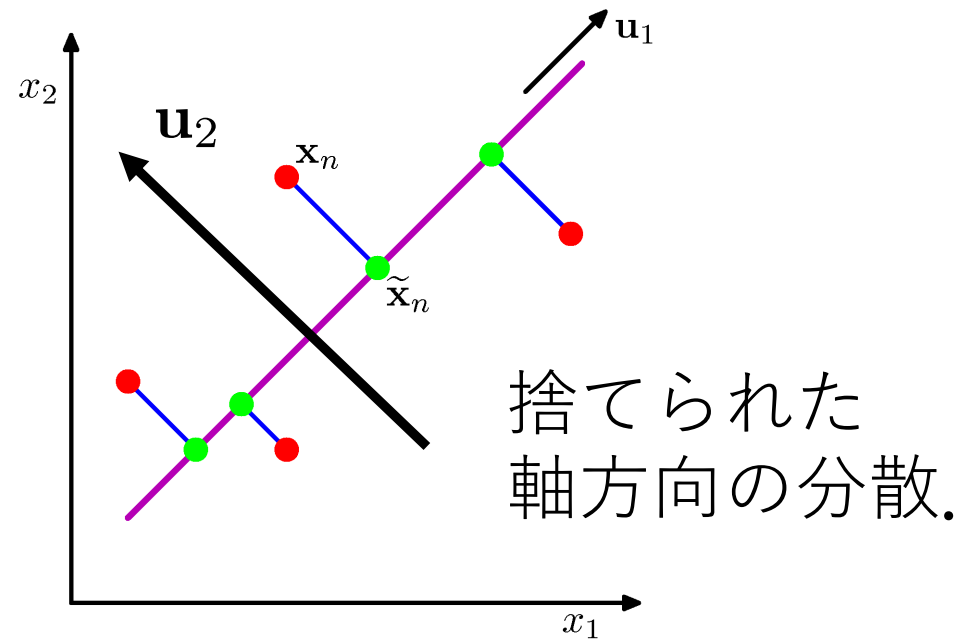
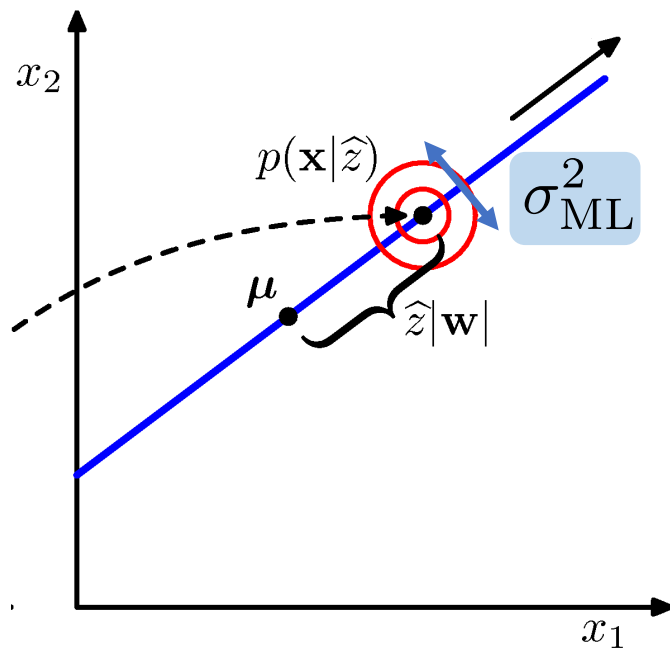
$$\mathbf{x} = \mathbf{W}_{\text{ML}} \mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

$$\lambda_i = \lambda_i - \sigma^2 + \sigma^2$$

後述

➤ 切り捨てられた次元に関する分散の平均.

$$\sigma_{\text{ML}}^2 = \frac{1}{D - M} \sum_{i=M+1}^D \lambda_i$$



$$\mathbf{W}_{\text{ML}} = \mathbf{U}_M (\mathbf{L}_M - \sigma_{\text{ML}}^2 \mathbf{I})^{1/2} \mathbf{R}$$

➤ 予測分布: $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$ $\mathbf{C} = \mathbf{W}\mathbf{W}^T + \sigma^2 \mathbf{I}$

$\mathbf{R}\mathbf{R}^T = \mathbf{I}$ であるから,

$$\mathbf{W}_{\text{ML}} \mathbf{W}_{\text{ML}}^T = \mathbf{U}_M (\mathbf{L}_M - \sigma_{\text{ML}}^2 \mathbf{I}) \mathbf{U}_M^T$$

- 潜在変数空間の回転に対し予測分布は影響を受けない.
- 同一の予測分布を与える \mathbf{W}_{ML} の族が存在する.

- 統計学的な識別不可能性.

識別不能 (しきべつふのう) とは、二つの確率変数を見分けることができないことを意味する。

- 離散潜在変数の混合モデル.

※wiki

- K個の混合要素に対してK!個の同等なモデルが存在.

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{z})p(\mathbf{x}|\mathbf{z}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

- PPCA.

- 任意の回転行列が同等の予測分布を与える.

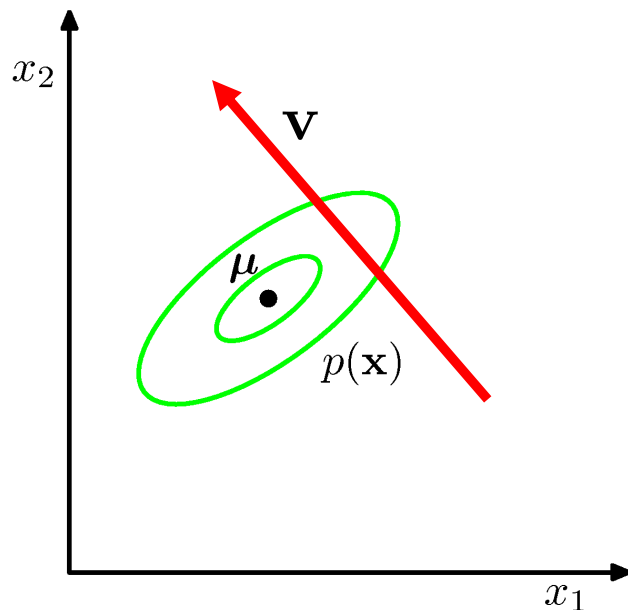
$$\tilde{\mathbf{W}} = \mathbf{W}\mathbf{R}$$

$$\tilde{\mathbf{W}}\tilde{\mathbf{W}}^T = \mathbf{W}\mathbf{R}\mathbf{R}^T\mathbf{W}^T = \mathbf{W}\mathbf{W}^T$$

- ある単位ベクトル \mathbf{v} に沿った予測分布の分散.
- \mathbf{v} が主部分空間に直交すると仮定.

$$\mathbf{v}^T \mathbf{v} = 1$$

$$\mathbf{v}^T \mathbf{U}_M = 0$$

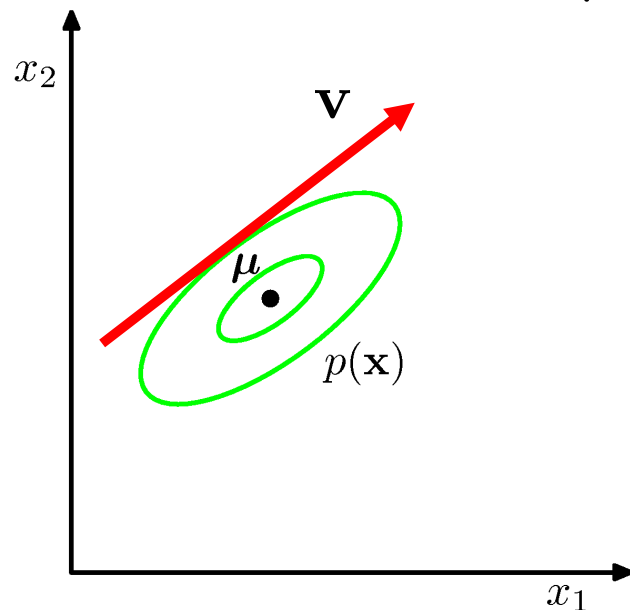


$$\begin{aligned} \mathbf{v}^T \mathbf{C} \mathbf{v} &= \mathbf{v}^T (\mathbf{U}_M (\mathbf{L}_M - \sigma_{\text{ML}}^2 \mathbf{I}) \mathbf{U}_M - \sigma_{\text{ML}}^2 \mathbf{I}) \mathbf{v} \\ &= 0 + \sigma_{\text{ML}}^2 \\ &= \frac{1}{D - M} \sum_{i=M+1}^D \lambda_i \end{aligned}$$

➤ \mathbf{v} が主部分空間を構成すると仮定.

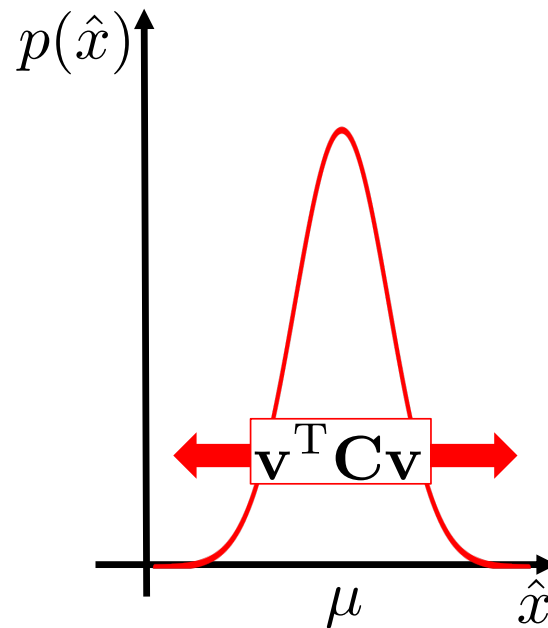
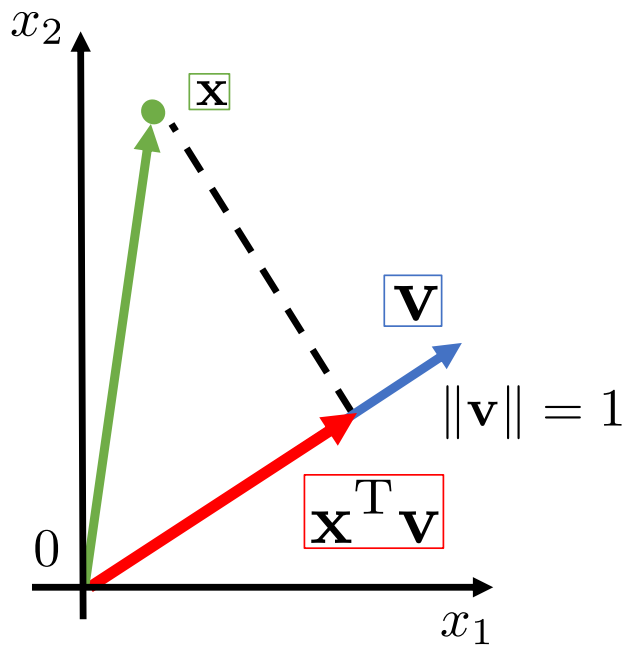
$$\mathbf{v} = \mathbf{u}_i$$

$$\begin{aligned}\mathbf{v}^T \mathbf{C} \mathbf{v} &= \mathbf{u}_i^T (\mathbf{U}_M (\mathbf{L}_M - \sigma_{ML}^2 \mathbf{I}) \mathbf{U}_M - \sigma_{ML}^2 \mathbf{I}) \mathbf{u}_i \\ &= (\lambda_i - \sigma_{ML}^2) + \sigma_{ML}^2 = \lambda_i\end{aligned}$$



- 主成分の軸方向の分散は正確に捉える.
- 残りの軸の分散は単一の値 σ_{ML}^2 で近似.

- 単位ベクトル \mathbf{v} 方向の予測分布の分散を新しい変数 $\mathbf{x}^T \mathbf{v}$ の従う分布における分散だと解釈すれば…,
- ガウス分布の再生性から,



$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{x}^T \mathbf{v} \\ \mathbf{x} &\sim \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}, \mathbf{C}) \end{aligned}$$



$$\hat{\mathbf{x}} \sim \mathcal{N}(\hat{\mathbf{x}} | \boldsymbol{\mu}, \mathbf{v}^T \mathbf{C} \mathbf{v})$$

➤ $M = D$ の場合,

$$\mathbf{L}_M = \mathbf{L}$$

$$\mathbf{U}\mathbf{U}^T = \mathbf{I}$$

$$\mathbf{U}_M = \mathbf{U}$$

$$\mathbf{R}\mathbf{R}^T = \mathbf{I}$$

$$\begin{aligned}\mathbf{C} &= \mathbf{U} (\mathbf{L} - \sigma^2 \mathbf{I})^{1/2} \mathbf{R}\mathbf{R}^T (\mathbf{L} - \sigma^2 \mathbf{I})^{1/2} \mathbf{U}^T + \sigma^2 \mathbf{I} \\ &= \mathbf{U}\mathbf{L}\mathbf{U}^T \\ &= \mathbf{S}\end{aligned}$$

➤ 制約のないガウス分布に対する標準的な最尤解.

- \mathbf{S} の固有値と固有ベクトルを求め、 \mathbf{W}_{ML} と σ_{ML}^2 を計算する。

終わり

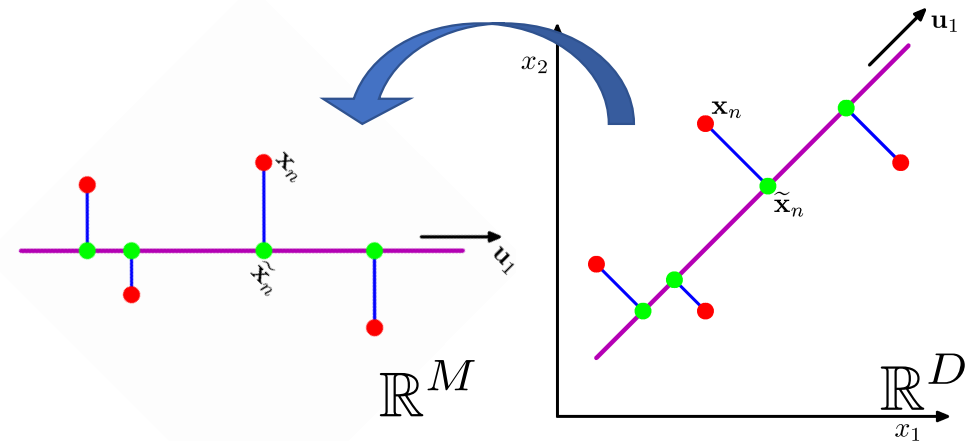
しかし、

- 共役勾配法, EMアルゴリズムで求めると,
 - 結果として \mathbf{R} の値は任意. (\mathbf{W} の列ベクトルが直交する必要なし)
- 直交基底を得るために, 後処理が可能.
- EMアルゴリズムを修正し, 固有値の大きい順に直交する主成分を直接得ることも可能.

➤ PCA.

- データ空間から線形部分空間への写像.

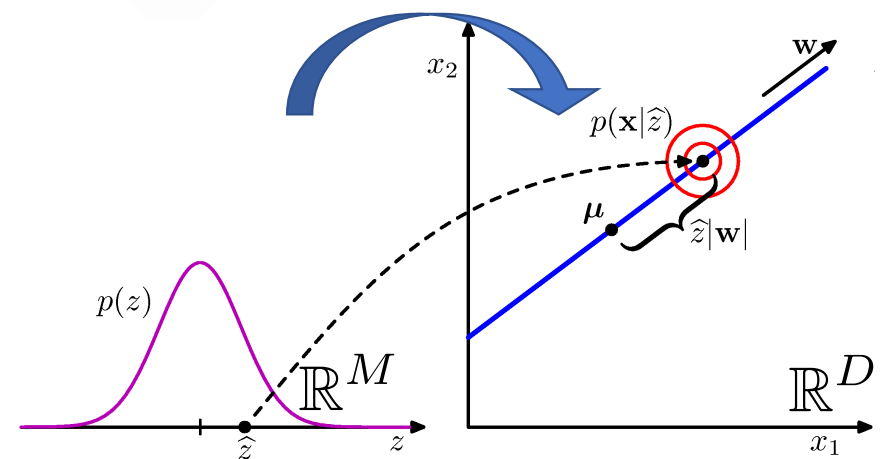
$$\mathbf{y}_n = \mathbf{L}^{-1/2} \mathbf{U}^T (\mathbf{x}_n - \bar{\mathbf{x}})$$



➤ PPCA.

- 潜在変数空間からデータ空間への写像.

$$\mathbf{x} = \mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon}$$



- 可視化, データ圧縮のために逆写像を得る.
- 事後分布.

$$p(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}|\mathbf{M}^{-1}\mathbf{W}^T(\mathbf{x} - \boldsymbol{\mu}), \sigma^2\mathbf{M}^{-1})$$

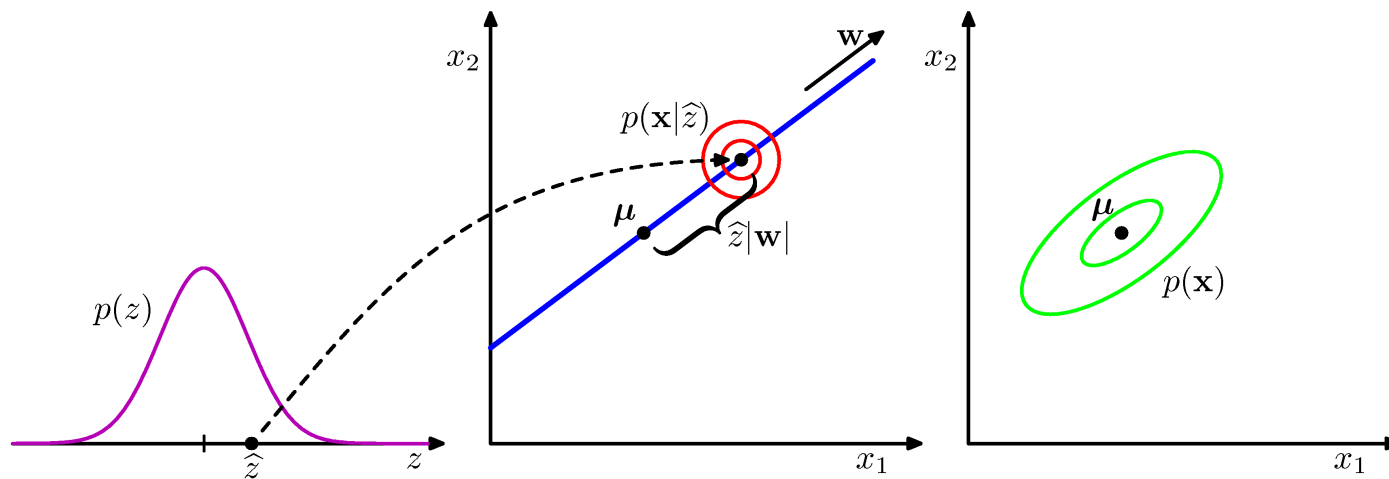
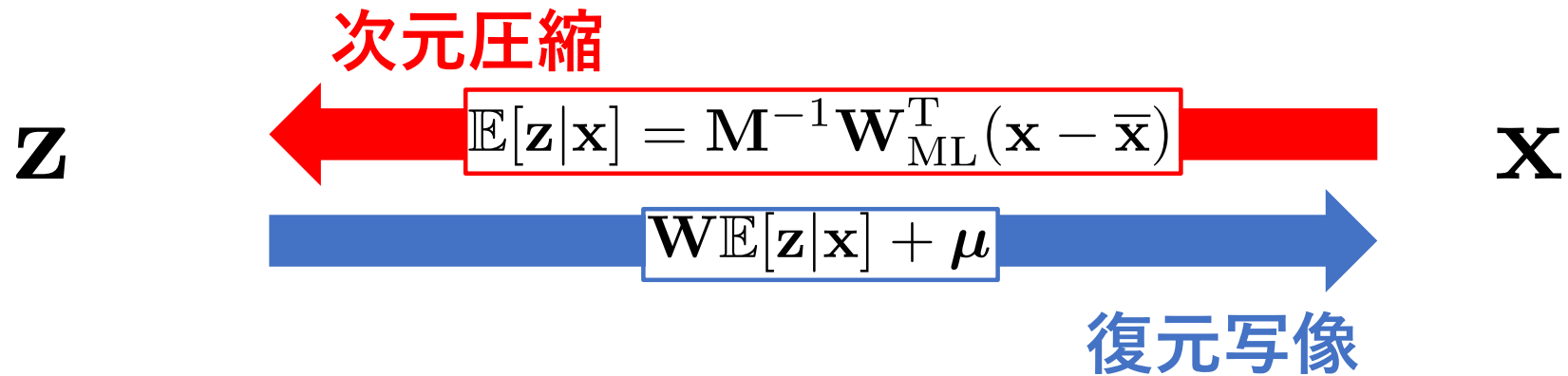
- 平均(次元圧縮).

$$\mathbb{E}[\mathbf{z}|\mathbf{x}] = \mathbf{M}^{-1}\mathbf{W}_{\text{ML}}^T(\mathbf{x} - \bar{\mathbf{x}})$$

$$\text{where } \mathbf{M} = \mathbf{W}_{\text{ML}}^T\mathbf{W}_{\text{ML}} + \sigma_{\text{ML}}^2\mathbf{I}$$

- 復元写像.

$$\mathbf{W}\mathbb{E}[\mathbf{z}|\mathbf{x}] + \boldsymbol{\mu}$$



- $\sigma_{\text{ML}}^2 \rightarrow 0$ の極限では,
 $\mathbf{R} = \mathbf{I}$ とすると

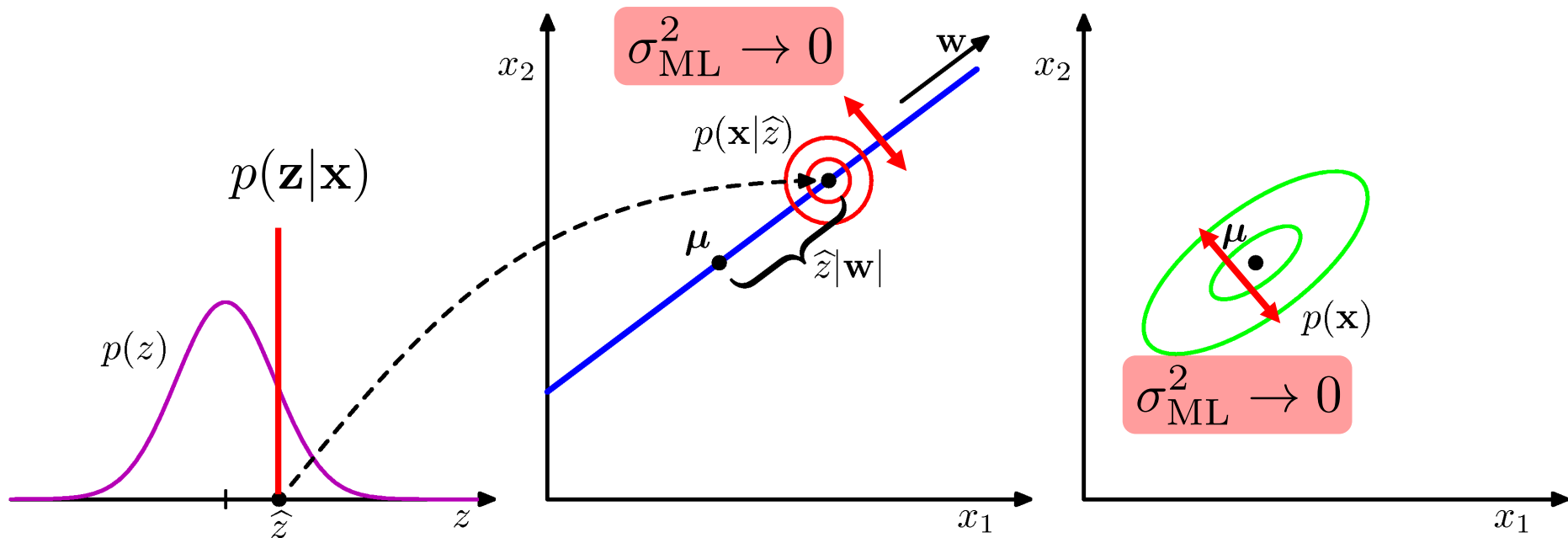
$$\begin{aligned} \mathbb{E}[\mathbf{z}|\mathbf{x}] &= \mathbf{M}^{-1} \mathbf{W}_{\text{ML}}^{\text{T}} (\mathbf{x} - \bar{\mathbf{x}}) \\ \text{where } \mathbf{M} &= \mathbf{W}_{\text{ML}}^{\text{T}} \mathbf{W}_{\text{ML}} + \sigma_{\text{ML}}^2 \mathbf{I} \\ \mathbf{W}_{\text{ML}} &= \mathbf{U}_M (\mathbf{L}_M - \sigma_{\text{ML}}^2 \mathbf{I})^{1/2} \mathbf{R} \end{aligned}$$

$$\begin{aligned} \mathbb{E}[\mathbf{z}|\mathbf{x}] &= (\mathbf{W}_{\text{ML}}^{\text{T}} \mathbf{W}_{\text{ML}})^{-1} \mathbf{W}_{\text{ML}}^{\text{T}} (\mathbf{x} - \bar{\mathbf{x}}) \\ &= \left(\left(\mathbf{U}_M \mathbf{L}_M^{1/2} \right)^{\text{T}} \left(\mathbf{U}_M \mathbf{L}_M^{1/2} \right) \right)^{-1} \left(\mathbf{U}_M \mathbf{L}_M^{1/2} \right)^{\text{T}} (\mathbf{x} - \bar{\mathbf{x}}) \\ &= \mathbf{L}_M^{-1/2} \mathbf{U}_M^{\text{T}} (\mathbf{x} - \bar{\mathbf{x}}) \end{aligned}$$

- データ点の潜在変数空間上への直交射影.
➤ 標準的な主成分分析モデルの再現.

➤ $\sigma_{\text{ML}}^2 \rightarrow 0$ の極限では事後分布は確率的でなくなる。

$$\text{COV} [\mathbf{z}|\mathbf{x}] = \sigma_{\text{ML}}^2 \mathbf{M}^{-1} = 0$$



➤ $\sigma_{\text{ML}}^2 > 0$ の場合,

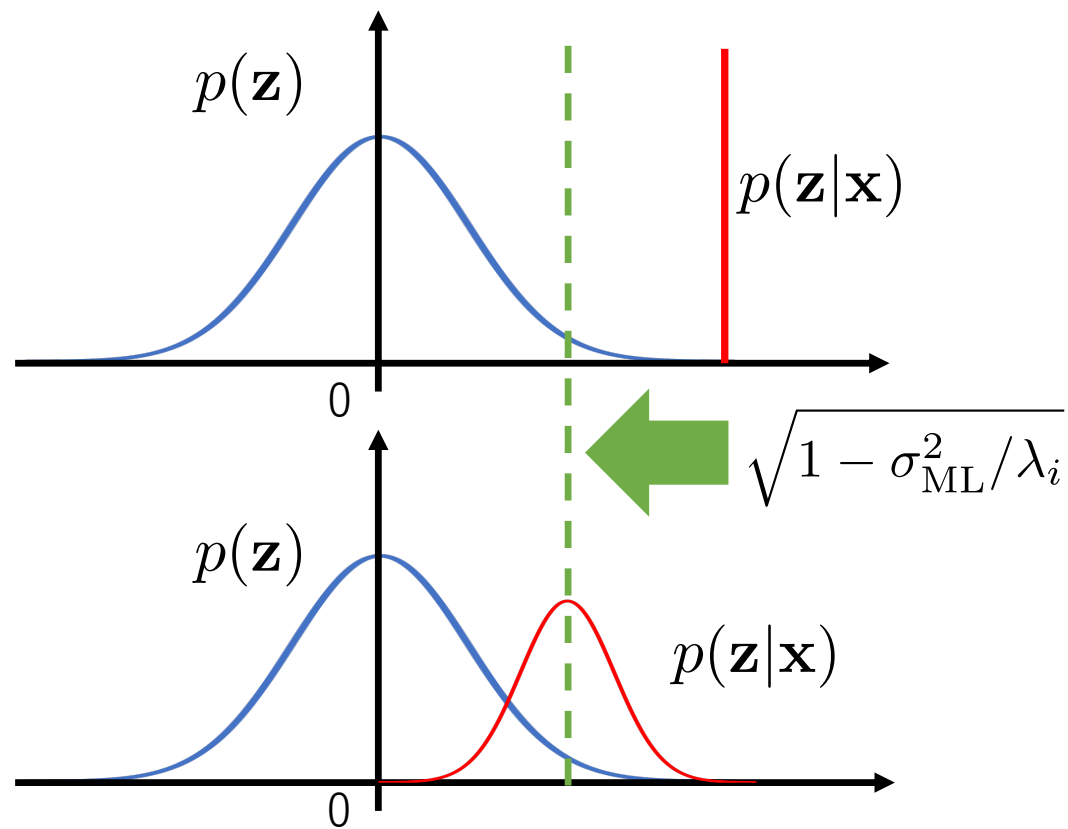
$$\begin{aligned} \mathbb{E}[\mathbf{z}|\mathbf{x}] &= \mathbf{M}^{-1} \mathbf{W}_{\text{ML}}^{\text{T}} (\mathbf{x} - \bar{\mathbf{x}}) \\ &= \mathbf{L}_M^{-1} (\mathbf{L}_M - \sigma_{\text{ML}}^2 \mathbf{I})^{1/2} \mathbf{U}_M^{\text{T}} (\mathbf{x} - \bar{\mathbf{x}}) \\ &= \mathbf{L}_M^{-1/2} (\mathbf{L}_M - \sigma_{\text{ML}}^2 \mathbf{I})^{1/2} \mathbf{L}_M^{-1/2} \mathbf{U}_M^{\text{T}} (\mathbf{x} - \bar{\mathbf{x}}) \\ &= (\mathbf{I} - \sigma_{\text{ML}}^2 \mathbf{L}^{-1})^{1/2} \mathbf{L}_M^{-1/2} \mathbf{U}_M^{\text{T}} (\mathbf{x} - \bar{\mathbf{x}}) \end{aligned}$$

$\sqrt{1 - \sigma_{\text{ML}}^2 / \lambda_i}$ スケーリング係数.

➤ $\sqrt{1 - \sigma_{\text{ML}}^2 / \lambda_i} < 1$ であるので, 原点側に射影点はシフトする.

$$\begin{aligned} \mathbf{M} &= \mathbf{W}_{\text{ML}}^{\text{T}} \mathbf{W}_{\text{ML}} + \sigma_{\text{ML}}^2 \mathbf{I} \\ &= \mathbf{L}_M - \sigma_{\text{ML}}^2 \mathbf{I} + \sigma_{\text{ML}}^2 \mathbf{I} \\ &= \mathbf{L}_M \end{aligned}$$

$$\mathbb{E}[\mathbf{z}|\mathbf{x}] = (\mathbf{I} - \sigma_{\text{ML}}^2 \mathbf{L}^{-1})^{1/2} \mathbf{L}_M^{-1/2} \mathbf{U}_M^T (\mathbf{x} - \bar{\mathbf{x}})$$



$$(\text{自由度}) = (\text{パラメータ数}) - (\text{制約式の数})$$

- 一般的なガウス分布
 - パラメータ数は D^2 に比例.
 - 共分散行列を対角に制限すれば, D に比例.
 - 重要な相関は無視.
- PPCA
 - パラメータ数が D に対して線形.
 - 主要な相関を M 個捉える.

- PPCA: $(DM + 1)$ 個

$$\mathbf{C} = \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I},$$
$$\mathbf{W} \in \mathbb{R}^{D \times M}$$

- 潜在変数空間の座標系の回転に関連した冗長性.

$$\mathbf{R}\mathbf{R}^T = \mathbf{I}$$

$$\mathbf{R} \in \mathbb{R}^{M \times M}$$

$M(M + 1)/2$ 個 の制約式

- \mathbf{R} の独立なパラメータは $M(M - 1)/2$ 個

- PPCAの共分散行列の自由度:

$$DM + 1 - M(M - 1)/2$$

- $M = D - 1$ の場合:

$D - 1$ 個数の分散を \mathbf{W} で, 残りを $\sigma^2 = \lambda_D$ で捉える.

- $M = D$ の場合:

$$\mathbf{C} = \mathbf{S}$$

- $M = 0$ の場合:

$$\sigma_{\text{ML}}^2 = \frac{1}{D - M} \sum_{i=M+1}^D \lambda_i$$



12.2.2 EMアルゴリズムによる主成分分析

- **潜在変数モデルの最尤解を求めるから.**
- 最尤解は閉形式で得られるが，高次元においては計算量的に有利.
- 因子分析では最尤解が閉形式でないが，EMによって計算可能.
- 欠損データが扱いやすい.

➤ Eステップ.

- 初期値 θ^{old} を決定し, 事後分布を計算.

$$p(\mathbf{Z}|\mathbf{X}, \theta^{\text{old}})$$

➤ Mステップ.

- 以下を計算.

$$\theta^{\text{new}} = \arg \max_{\theta} Q(\theta, \theta^{\text{old}})$$

where $Q(\theta, \theta^{\text{old}}) = \int p(\mathbf{Z}|\mathbf{X}, \theta^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\theta) d\mathbf{Z}$

完全データ対数尤度関数

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \int p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) d\mathbf{Z}$$

where

$$\boldsymbol{\theta} = \{\boldsymbol{\mu}, \mathbf{W}, \sigma^2\}$$

$$\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^N \quad \mathbf{x}_n | \mathbf{z}_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mathbf{W}\mathbf{z}_n + \boldsymbol{\mu}, \sigma^2 \mathbf{I})$$

$$\mathbf{Z} = \{\mathbf{z}_n\}_{n=1}^N \quad \mathbf{z}_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$\ln p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \mathbf{W}, \sigma^2) = \sum_{n=1}^N \{\ln p(\mathbf{x}_n | \mathbf{z}_n) + \ln p(\mathbf{z}_n)\}$$

$$\mathbb{E}_{\mathbf{z}|\mathbf{x}} \left[\sum_{n=1}^N \ln p(\mathbf{x}_n | \mathbf{z}_n) \right]$$

$$\begin{aligned} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) &= \mathbb{E}_{\mathbf{z}|\mathbf{x}} [\ln p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \mathbf{W}, \sigma^2)] \\ &= \mathbb{E}_{\mathbf{z}|\mathbf{x}} \left[\sum_{n=1}^N \{ \ln p(\mathbf{x}_n | \mathbf{z}_n) + \ln p(\mathbf{z}_n) \} \right] \end{aligned}$$

$$= \sum_{n=1}^N \left\{ -\frac{D}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \mathbb{E}_{\mathbf{z}|\mathbf{x}} \left[(\mathbf{x}_n - \mathbf{W}\mathbf{z}_n - \boldsymbol{\mu})^T (\mathbf{x}_n - \mathbf{W}\mathbf{z}_n - \boldsymbol{\mu}) \right] \right\}$$

$$= \sum_{n=1}^N \left\{ -\frac{D}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \mathbb{E}_{\mathbf{z}|\mathbf{x}} \left[(\mathbf{x}_n - \boldsymbol{\mu})^T (\mathbf{x}_n - \boldsymbol{\mu}) - 2(\mathbf{W}\mathbf{z}_n)^T (\mathbf{x}_n - \boldsymbol{\mu}) + (\mathbf{W}\mathbf{z}_n)^T (\mathbf{W}\mathbf{z}_n) \right] \right\}$$

$$= \sum_{n=1}^N \left\{ -\frac{D}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left(\|\mathbf{x}_n - \boldsymbol{\mu}\|^2 - 2\mathbb{E}_{\mathbf{z}|\mathbf{x}} [\mathbf{z}_n^T] \mathbf{W}^T (\mathbf{x}_n - \boldsymbol{\mu}) + \text{Tr}(\mathbb{E}_{\mathbf{z}|\mathbf{x}} [\mathbf{z}_n \mathbf{z}_n^T] \mathbf{W}^T \mathbf{W}) \right) \right\}$$

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) &= \mathbb{E}_{\mathbf{z}|\mathbf{x}} [\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \mathbf{W}, \sigma^2)] \\ &= \mathbb{E}_{\mathbf{z}|\mathbf{x}} \left[\sum_{n=1}^N \{ \ln p(\mathbf{x}_n|\mathbf{z}_n) + \ln p(\mathbf{z}_n) \} \right] \end{aligned}$$

$$\begin{aligned} \mathbb{E}_{\mathbf{z}|\mathbf{x}} \left[\sum_{n=1}^N \ln p(\mathbf{z}_n) \right] &= \sum_{n=1}^N \left\{ -\frac{M}{2} \ln(2\pi) - \frac{1}{2} \mathbb{E}_{\mathbf{z}|\mathbf{x}} [\mathbf{z}_n^T \mathbf{I} \mathbf{z}_n] \right\} \\ &= \sum_{n=1}^N \left\{ -\frac{M}{2} \ln(2\pi) - \frac{1}{2} \text{Tr}(\mathbb{E}_{\mathbf{z}|\mathbf{x}} [\mathbf{z}_n \mathbf{z}_n^T]) \right\} \end{aligned}$$

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) &= \mathbb{E}_{\mathbf{z}|\mathbf{x}} [\ln p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \mathbf{W}, \sigma^2)] \\ &= - \sum_{n=1}^N \left\{ \frac{D}{2} \ln(2\pi\sigma^2) + \frac{1}{2} \text{Tr}(\mathbb{E}_{\mathbf{z}|\mathbf{x}}[\mathbf{z}_n \mathbf{z}_n^T]) \right. \\ &\quad + \frac{1}{2\sigma^2} \|\mathbf{x}_n - \boldsymbol{\mu}\|^2 - \frac{1}{\sigma^2} \mathbb{E}_{\mathbf{z}|\mathbf{x}}[\mathbf{z}_n]^T \mathbf{W}^T (\mathbf{x}_n - \boldsymbol{\mu}) \\ &\quad \left. + \frac{1}{2\sigma^2} \text{Tr}(\mathbb{E}_{\mathbf{z}|\mathbf{x}}[\mathbf{z}_n \mathbf{z}_n^T] \mathbf{W}^T \mathbf{W}) + \frac{M}{2} \ln(2\pi) \right\} \end{aligned}$$

- 初期値を決めて事後分布(十分統計量)を計算.

$$p(\mathbf{z}_n | \mathbf{x}_n) = \mathcal{N}(\mathbf{z}_n | \mathbf{M}^{-1} \mathbf{W}^T (\mathbf{x}_n - \boldsymbol{\mu}), \sigma^2 \mathbf{M}^{-1})$$

$$\text{where } \mathbf{M} = \mathbf{W}^T \mathbf{W} + \sigma^2 \mathbf{I}$$

- 統計量の計算.

$$\mathbb{E}_{\mathbf{z} | \mathbf{x}} [\mathbf{z}_n] = \mathbf{M}^{-1} \mathbf{W}^T (\mathbf{x}_n - \bar{\mathbf{x}})$$

$$\mathbb{E}_{\mathbf{z} | \mathbf{x}} [\mathbf{z}_n \mathbf{z}_n^T] = \sigma^2 \mathbf{M}^{-1} + \mathbb{E}_{\mathbf{z} | \mathbf{x}} [\mathbf{z}_n] \mathbb{E}_{\mathbf{z} | \mathbf{x}} [\mathbf{z}_n]^T$$

➤ Eステップで得た統計量を固定し以下でパラメータ更新.

$$\mathbf{W}_{\text{new}} = \left[\sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}}) \mathbb{E}_{\mathbf{z}|\mathbf{x}} [\mathbf{z}_n]^T \right] \left[\sum_{n=1}^N \mathbb{E}_{\mathbf{z}|\mathbf{x}} [\mathbf{z}_n \mathbf{z}_n^T] \right]^{-1}$$

$$\begin{aligned} \sigma_{\text{new}}^2 = & \frac{1}{ND} \sum_{n=1}^N \left\{ \|\mathbf{x}_n - \bar{\mathbf{x}}\|^2 - 2 \mathbb{E}_{\mathbf{z}|\mathbf{x}} [\mathbf{z}_n]^T \mathbf{W}_{\text{new}}^T (\mathbf{x}_n - \bar{\mathbf{x}}) \right. \\ & \left. + \text{Tr} \left(\mathbb{E}_{\mathbf{z}|\mathbf{x}} [\mathbf{z}_n \mathbf{z}_n^T] \mathbf{W}_{\text{new}}^T \mathbf{W}_{\text{new}} \right) \right\} \end{aligned}$$

- 通常のPCA
 - 共分散行列の計算: $O(ND^2)$
 - 共分散行列の固有値分解: $O(D^3)$
 - M個の主成分を求める: $O(MD^2)$
- スナップショット法
 - 固有ベクトルをデータベクトルの一次結合と仮定.
 - 共分散行列の計算は避けられるが, $O(N^3)$
- EMアルゴリズム
 - データ集合にわたる和: $O(NDM)$
 - **高次元データに有利.**

➤ Eステップ

- データ点ごとに以下の統計量を計算.

$$\mathbb{E}_{\mathbf{z}|\mathbf{x}} [\mathbf{z}_n], \quad \mathbb{E}_{\mathbf{z}|\mathbf{x}} [\mathbf{z}_n \mathbf{z}_n^T]$$

➤ Mステップ

- データ集合にわたる和の計算.
 - 増加分だけ計算すれば良い.
-
- データ集合が大きく、高次元の場合に有利.

- 欠損データを非観測変数として扱う.
- 同時分布から欠損値について周辺化(9.3節参考).

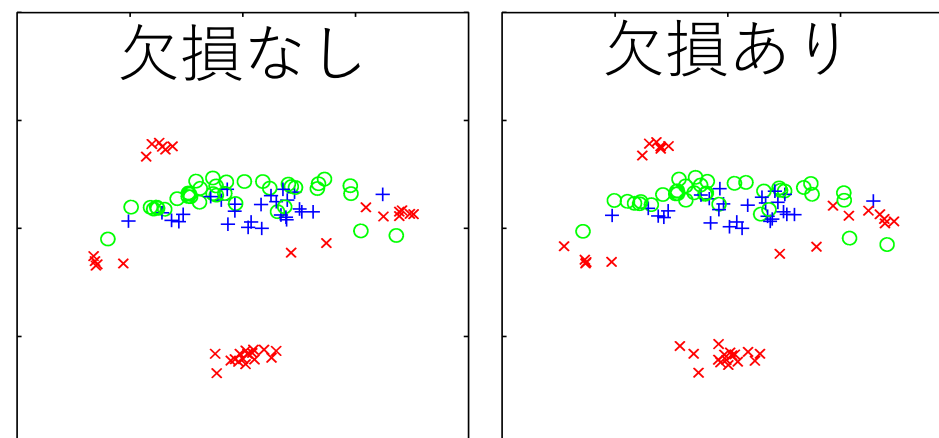
$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \int p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) d\mathbf{Z}$$

- 送油データの例.

$$\mathbf{X} = \{\mathbf{x}_n\}_{n=1}^{100}, \quad \mathbf{x}_n \in \mathbb{R}^{12}$$

- 変数値の30%がランダム欠損.
- パイプ内の相配置に関するラベル数が3.

事後平均 $\mathbb{E}_{\mathbf{z}|\mathbf{x}}[\mathbf{z}_n]$ のプロット.



$\sigma^2 \rightarrow 0$ の場合におけるEM的アルゴリズムを導出できる.

➤ Eステップ

➤ 必要な統計量は $\mathbb{E}_{\mathbf{z}|\mathbf{x}}[\mathbf{z}_n]$ だけ.

$$\mathbf{\Omega} = (\mathbf{W}_{\text{old}}^T \mathbf{W}_{\text{old}})^{-1} \mathbf{W}_{\text{old}}^T \tilde{\mathbf{X}}^T$$

$\mathbf{\Omega}$: 列ベクトルが $\mathbb{E}_{\mathbf{z}|\mathbf{x}}[\mathbf{z}_n]$ で与えられる $M \times N$ 行列.

$\tilde{\mathbf{X}}$: 行ベクトルが $(\mathbf{x}_n - \bar{\mathbf{x}})^T$ で与えられる $N \times D$ 行列.

➤ Mステップ

$$\mathbf{W}_{\text{new}} = \tilde{\mathbf{X}}^T \mathbf{\Omega}^T (\mathbf{\Omega} \mathbf{\Omega}^T)^{-1}$$

- Eステップ
 - 現時点で推定される主部分空間へのデータ点の直交射影.
- Mステップ
 - 射影先を固定した2乗再現誤差の最小化による主部分空間の再推定.

$$J = \sum_{n=1}^N \|\mathbf{x}_n - \boldsymbol{\mu} - \mathbf{W}\mathbf{z}_n\|^2$$

