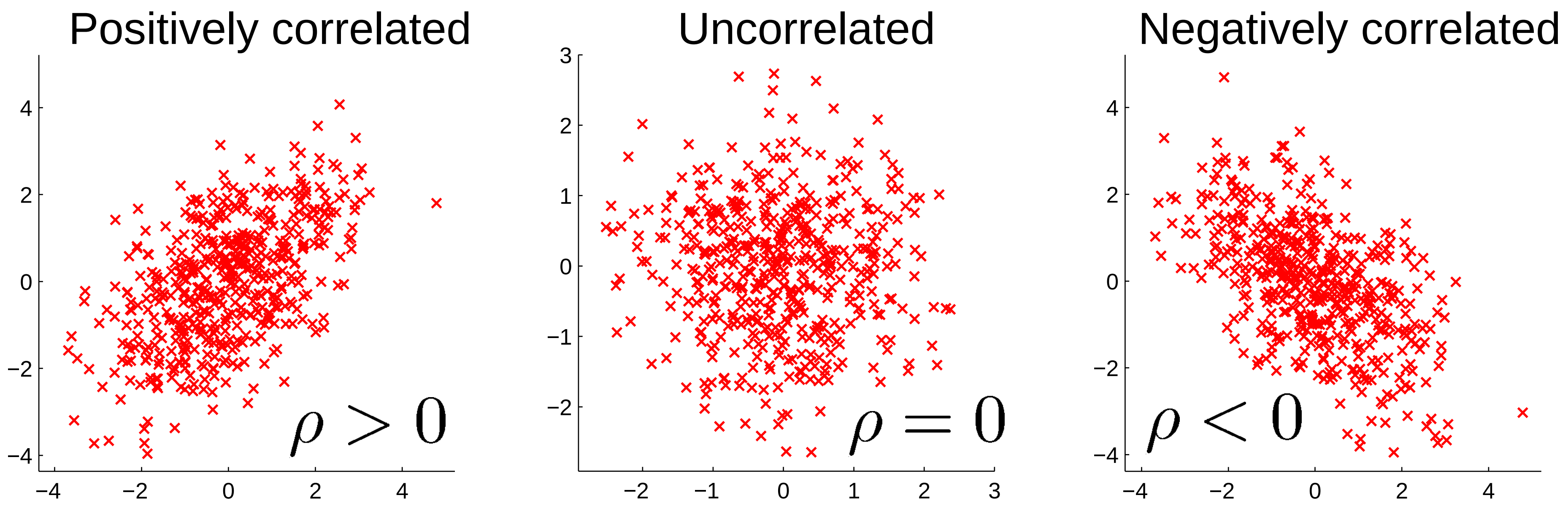
Correlation

- Correlation coefficient for $\{s_i, t_i\}_{i=1}^n$:

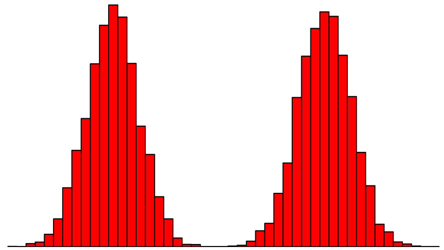
$$\rho = \frac{\sum_{i=1}^n (s_i - \overline{s})(t_i - \overline{t})}{\sqrt{\left(\sum_{i=1}^n (s_i - \overline{s})^2\right)\left(\sum_{i=1}^n (t_i - \overline{t})^2\right)}}$$

Correlation



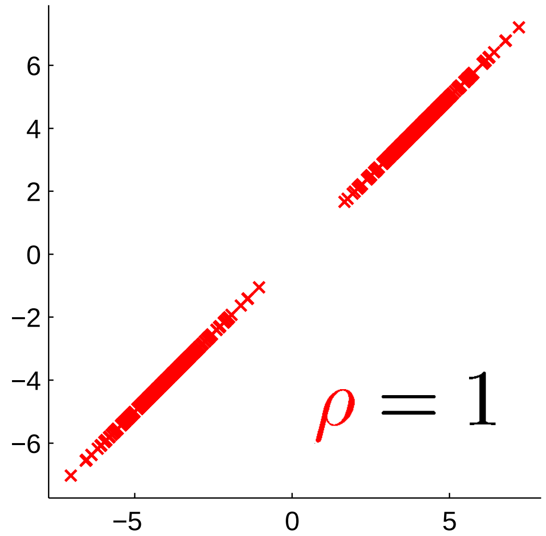
Redundancy

- We observe the following one-dimensional signal.



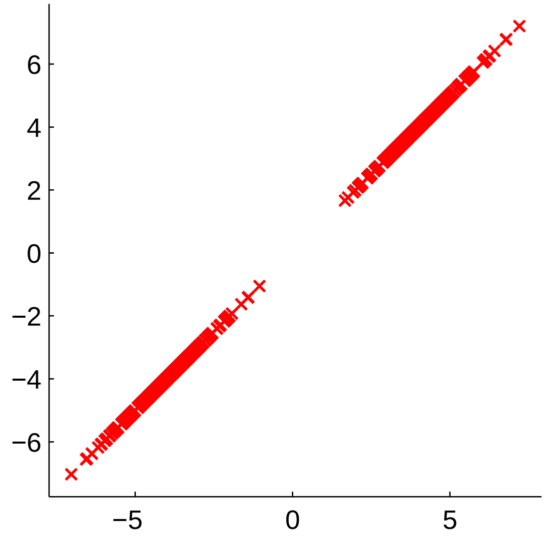
- If two devices observe exactly the same signal, data has (maximum) positive correlation: $\rho = 1$

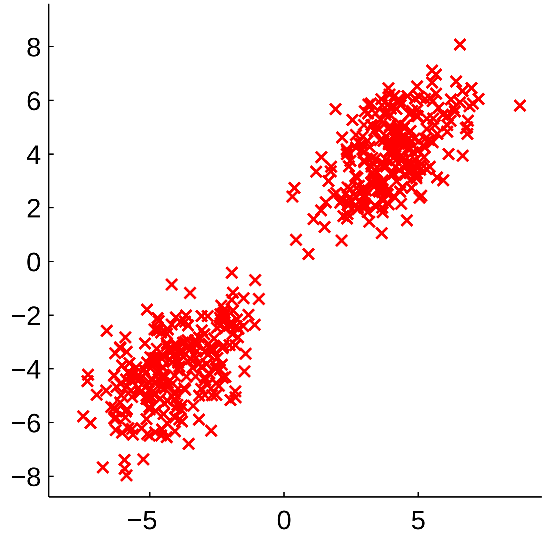
Observed signal is redundant



High-Dimensional Data

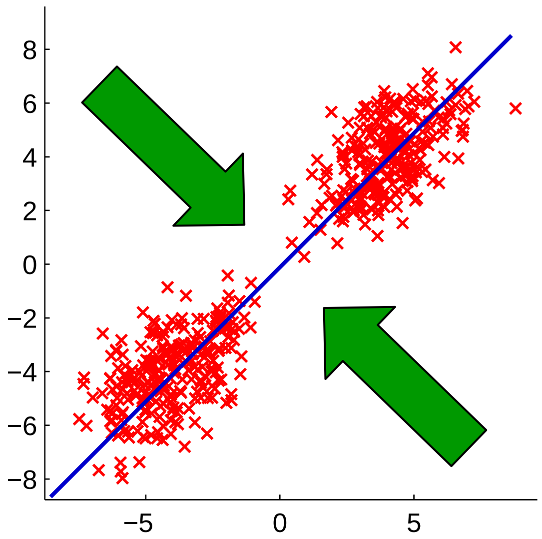
- However, in practical data with high-dimension, attributes are often redundant and correlated.
- Furthermore, they are often degraded by noise.





Noise Reduction

- We want to reduce the influence of noise and extract cleaner signals.
- This can be achieved by projecting out noisy directions without signals.



Noise Reduction (cont.)

- Suppose
  - Signals are non-Gaussian
  - Noises are Gaussian
- Under this assumption, noise can be reduced by projecting out directions without non-Gaussian components.
- Projection pursuit may be used for this purpose.

Non-Gaussianity Measures

- PP needs to pre-specify non-Gaussianity measure.
- It is known that
  - Some NG measures (e.g., log-cosh) are suitable for finding super-Gaussian components.
  - Others (e.g., kurtosis) are suitable for finding sub-Gaussian components.
- Here, we give another approach to finding non-Gaussian components.

Data Density

- Observation
  = Redundant non-Gaussian signal
  + Gaussian noise

$$\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{\epsilon}$$

$$\boldsymbol{s} \sim q(\boldsymbol{s}) \qquad \boldsymbol{\epsilon} \sim \phi_{\boldsymbol{\theta},\boldsymbol{\Gamma}}(\boldsymbol{\epsilon})$$

- Density function $p(\boldsymbol{x})$ is given by the convolution of two density functions:

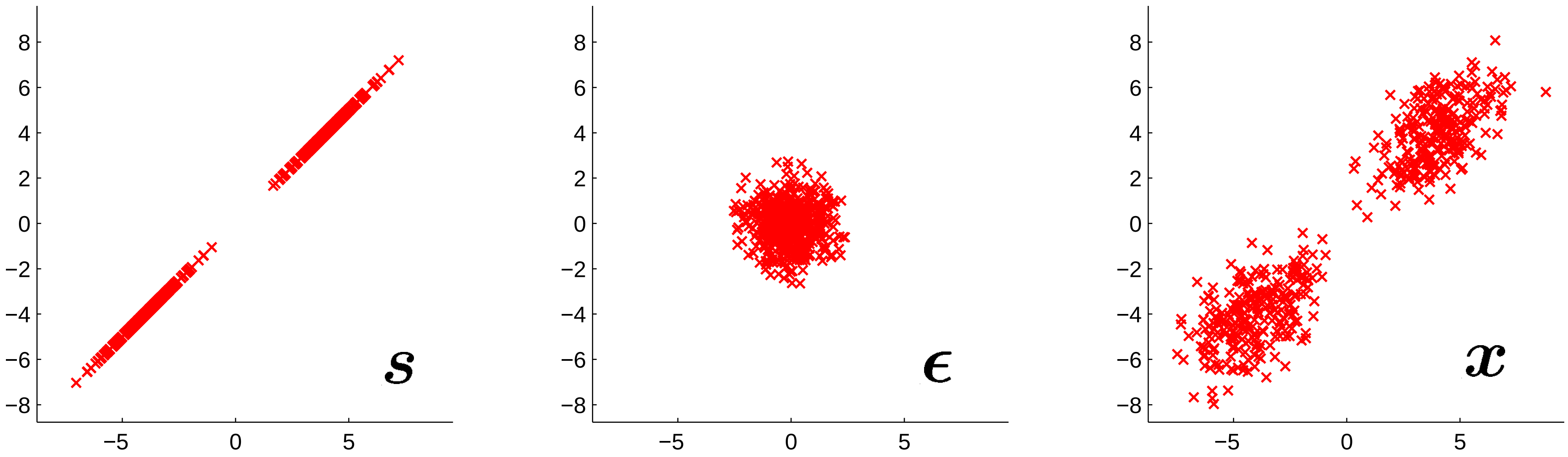
$$p(\boldsymbol{x}) = \int q(\boldsymbol{s})\phi_{\boldsymbol{\theta},\boldsymbol{\Gamma}}(\boldsymbol{x} - \boldsymbol{s})d\boldsymbol{s}$$

Data Density (cont.)

$$p(\boldsymbol{x}) = \int q(\boldsymbol{s})\phi_{\boldsymbol{\theta}+\boldsymbol{s},\boldsymbol{\Gamma}}(\boldsymbol{x})d\boldsymbol{s}$$

- Intuition: "Sum of many Gaussians"

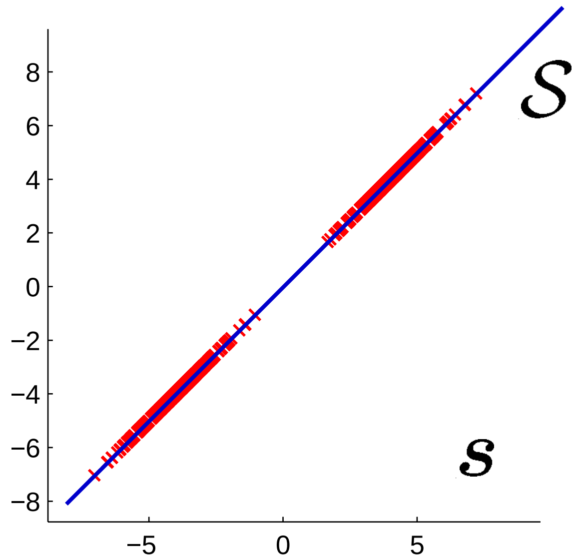
$$\boldsymbol{x} = \boldsymbol{s} + \boldsymbol{\epsilon}$$



Data Density (cont.)

- Since $\boldsymbol{s}$ is redundant, $q(\boldsymbol{s}) > 0$ only in some subspace $\mathcal{S}$.

$$p(\boldsymbol{x}) = \int_{\boldsymbol{s} \in \mathcal{S}} q(\boldsymbol{s}) \phi_{\boldsymbol{\theta}, \boldsymbol{\Gamma}}(\boldsymbol{x} - \boldsymbol{s}) d\boldsymbol{s}$$



Data Density (cont.)

- Suppose that Gaussian noise is mean zero and covariance identity:

$$\phi(\boldsymbol{\epsilon}) = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}\|\boldsymbol{\epsilon}\|^2}$$

- Then $p(\boldsymbol{x})$ is expressed as

$$p(\boldsymbol{x}) = g(\boldsymbol{x})\phi(\boldsymbol{x})$$

$$g(\boldsymbol{x}) = \int_{\boldsymbol{s}\in\mathcal{S}} q(\boldsymbol{s}) e^{-\frac{1}{2}\|\boldsymbol{s}\|^2} e^{\langle \boldsymbol{s}, \boldsymbol{x}\rangle} d\boldsymbol{s}$$

Data Density (cont.)

$$g(\boldsymbol{x}) = \int_{\boldsymbol{s}\in\mathcal{S}} q(\boldsymbol{s}) e^{-\frac{1}{2}\|\boldsymbol{s}\|^2} e^{\langle \boldsymbol{s},\boldsymbol{x}\rangle} d\boldsymbol{s}$$

- $\boldsymbol{P}$:Projection onto $\mathcal{S}$
- $\boldsymbol{P}$ is expressed using a $m \times d$ matrix $\boldsymbol{T}$ as

$$\boldsymbol{P} = \boldsymbol{T}^\top \boldsymbol{T}$$

$$m = \dim(\mathcal{S})$$

- For $\boldsymbol{s} \in \mathcal{S}$,

$$\langle \boldsymbol{s},\boldsymbol{x}\rangle = \langle \boldsymbol{P}\boldsymbol{s},\boldsymbol{x}\rangle = \langle \boldsymbol{T}^\top\boldsymbol{T}\boldsymbol{s},\boldsymbol{x}\rangle = \langle \boldsymbol{s},\boldsymbol{T}^\top\boldsymbol{T}\boldsymbol{x}\rangle$$

$$\Rightarrow \quad p(\boldsymbol{x}) = f(\boldsymbol{T}\boldsymbol{x})\phi(\boldsymbol{x})$$

$$f(\boldsymbol{z}) = g(\boldsymbol{T}^\top\boldsymbol{z})$$

Non-Gaussian Subspace

- Orthogonal decomposition:

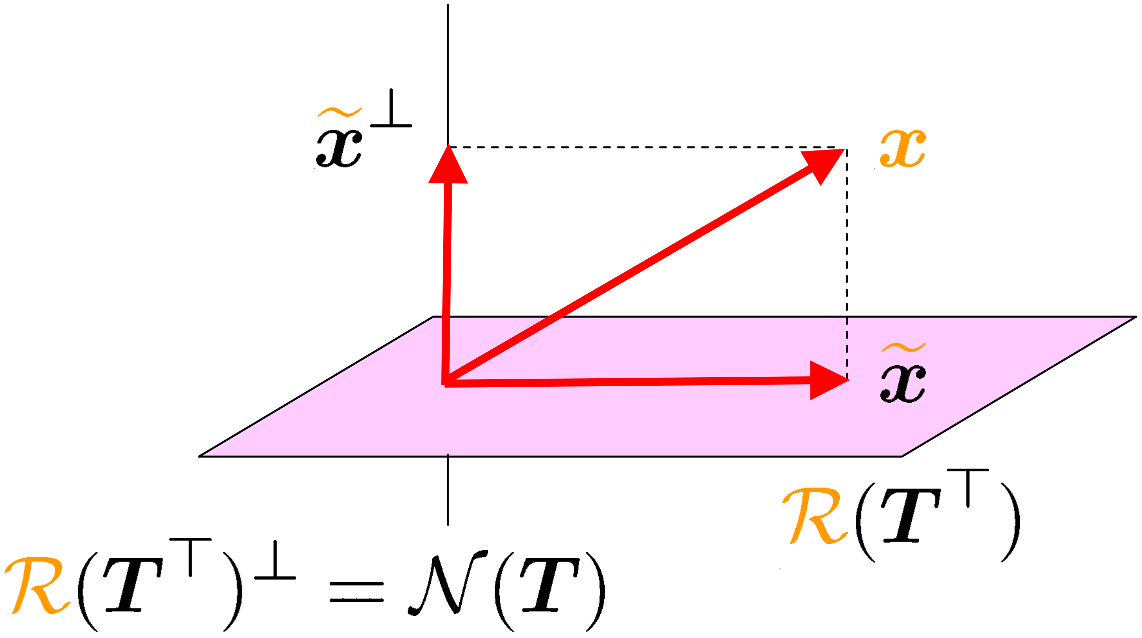
$$\boldsymbol{x} = \widetilde{\boldsymbol{x}} + \widetilde{\boldsymbol{x}}^{\perp}$$

$$\widetilde{\boldsymbol{x}} = \boldsymbol{P}\boldsymbol{x}$$

$$\widetilde{\boldsymbol{x}}^{\perp} = \boldsymbol{x} - \widetilde{\boldsymbol{x}}$$

- $\widetilde{\boldsymbol{x}} \in \mathcal{R}(\boldsymbol{T}^{\top})$
- $\widetilde{\boldsymbol{x}}^{\perp} \in \mathcal{R}(\boldsymbol{T}^{\top})^{\perp} = \mathcal{N}(\boldsymbol{T})$

$$\boldsymbol{P} = \boldsymbol{T}^{\top}\boldsymbol{T}$$

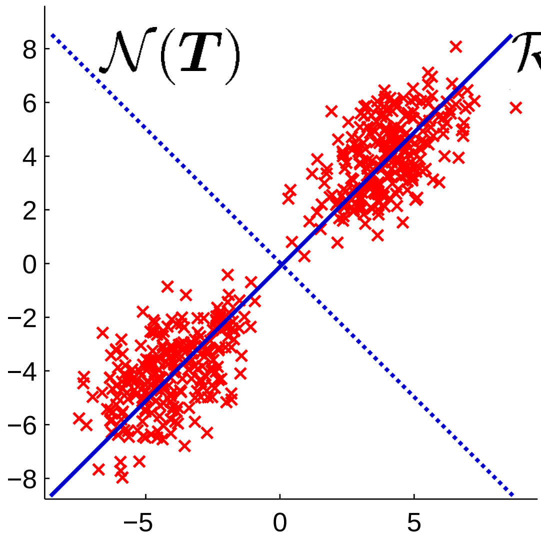


Non-Gaussian Subspace (cont.)

$$p(\boldsymbol{x}) = f(\boldsymbol{T}(\widetilde{\boldsymbol{x}} + \widetilde{\boldsymbol{x}}^{\perp}))\phi(\widetilde{\boldsymbol{x}} + \widetilde{\boldsymbol{x}}^{\perp})$$
$$= f(\boldsymbol{T}\widetilde{\boldsymbol{x}})\phi(\widetilde{\boldsymbol{x}})\phi(\widetilde{\boldsymbol{x}}^{\perp})$$

$$\widetilde{\boldsymbol{x}} \in \mathcal{R}(\boldsymbol{T}^{\top})$$
$$\widetilde{\boldsymbol{x}}^{\perp} \in \mathcal{N}(\boldsymbol{T})$$

- Given $\widetilde{\boldsymbol{x}}$, $\widetilde{\boldsymbol{x}}^{\perp}$ is conditionally Gaussian so it does not contain non-Gaussian components.
- $\mathcal{R}(\boldsymbol{T}^{\top})$:Non-Gaussian subspace
- We want to identify $\mathcal{R}(\boldsymbol{T}^{\top})$



General Framework

- Suppose $p(\boldsymbol{x})$ is expressed as

$$p(\boldsymbol{x}) = f(\boldsymbol{T}\boldsymbol{x})\phi_{\boldsymbol{\theta},\boldsymbol{\Gamma}}(\boldsymbol{x})$$

  - $\boldsymbol{T}$: Unknown linear mapping $\mathbb{R}^d \rightarrow \mathbb{R}^m$
  - f : Unknown function on $\mathbb{R}^m$
  - $\boldsymbol{\theta}, \boldsymbol{\Gamma}$: Unknown mean and covariance matrix

- We want to identify the non-Gaussian subspace $\mathcal{R}(\boldsymbol{T}^\top)$, without estimating

$$\boldsymbol{T}, f, \boldsymbol{\theta}, \boldsymbol{\Gamma}$$

- For simplicity, we assume

$$\int \boldsymbol{x}p(\boldsymbol{x}) = 0, \quad \int \boldsymbol{x}\boldsymbol{x}^\top p(\boldsymbol{x}) = \boldsymbol{I}_d, \quad \boldsymbol{\theta} = \boldsymbol{0}$$

Key Theorem

- $\psi(\boldsymbol{x})$: Smooth function from $\mathbb{R}^d$ to $\mathbb{R}$ which fulfills

$$\int \boldsymbol{x}\psi(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} = \mathbf{0} \quad \text{(A)}$$

- Then the vector $\boldsymbol{\beta}$ defined by

$$\boldsymbol{\beta}(\psi) = \int \nabla\psi(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$$

belongs to the non-Gaussian space $\mathcal{R}(\boldsymbol{T}^\top)$:

$$\boldsymbol{\beta}(\psi) \in \mathcal{R}(\boldsymbol{T}^\top)$$

Proof

$$\int \psi(\boldsymbol{x}+\boldsymbol{u})p(\boldsymbol{x})d\boldsymbol{x} = \int \psi(\boldsymbol{x})p(\boldsymbol{x}-\boldsymbol{u})d\boldsymbol{x}$$

- Differentiating this with respect to $\boldsymbol{u}$ gives

$$\int \nabla\psi(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} = -\int \psi(\boldsymbol{x})\nabla p(\boldsymbol{x})d\boldsymbol{x}$$

- Since $\nabla \log p(\boldsymbol{x}) = \dfrac{\nabla p(\boldsymbol{x})}{p(\boldsymbol{x})}$,

$$\boldsymbol{\beta}(\psi) = -\int \psi(\boldsymbol{x})\nabla \log p(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$$

Proof (cont.)

$$p(\boldsymbol{x}) = f(\boldsymbol{T}\boldsymbol{x})\phi_{\boldsymbol{\theta},\boldsymbol{\Gamma}}(\boldsymbol{x})$$

$$\begin{aligned}\nabla \log p(\boldsymbol{x}) &= \nabla \log f(\boldsymbol{T}\boldsymbol{x}) + \nabla \log \phi_{\boldsymbol{0},\boldsymbol{\Gamma}}(\boldsymbol{x}) \\ &= \frac{\boldsymbol{T}^\top \nabla f(\boldsymbol{T}\boldsymbol{x})}{f(\boldsymbol{T}\boldsymbol{x})} - \boldsymbol{\Gamma}^{-1}\boldsymbol{x}\end{aligned}$$

- This yields

$$\begin{aligned}\boldsymbol{\beta}(\psi) = -\boldsymbol{T}^\top \int \nabla f(\boldsymbol{T}\boldsymbol{x})\psi(\boldsymbol{x})\phi_{\boldsymbol{0},\boldsymbol{\Gamma}}(\boldsymbol{x})d\boldsymbol{x} \quad & \in \mathcal{R}(\boldsymbol{T}^\top) \\ +\boldsymbol{\Gamma}^{-1}\int \boldsymbol{x}\psi(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} \quad & = \boldsymbol{0}\end{aligned}$$

Homework

- Intermediate evaluation of this course:
  - Write your opinion about this course, e.g., contents, relevance, level, homework, explanation clarity, lecturer…
  - What did you like/dislike so far?
- I want to make this course interesting and useful for all of you!
- Do this homework seriously to improve the latter half of the course.