Dual Form, Duality Theorem, Lagrange multiplier and KKT,

Yan PEI/裴岩 /ペイ　イエン

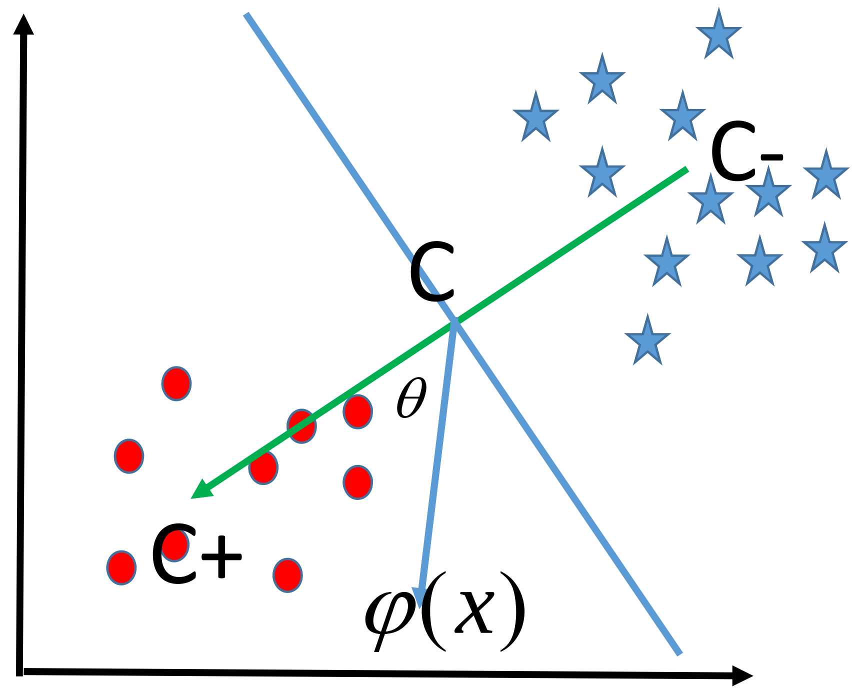
peiyan@u-aizu.ac.jp

Wahba Representer Theorem

The solutions of certain risk minimization problems involving an empirical risk term and a quadratic regularizer can be written as expansions in terms of the training samples.

$$f(x) = w^T \varphi(x) + b$$

$$w = \sum_{i=1}^{N} \alpha_i \varphi(x_i)$$

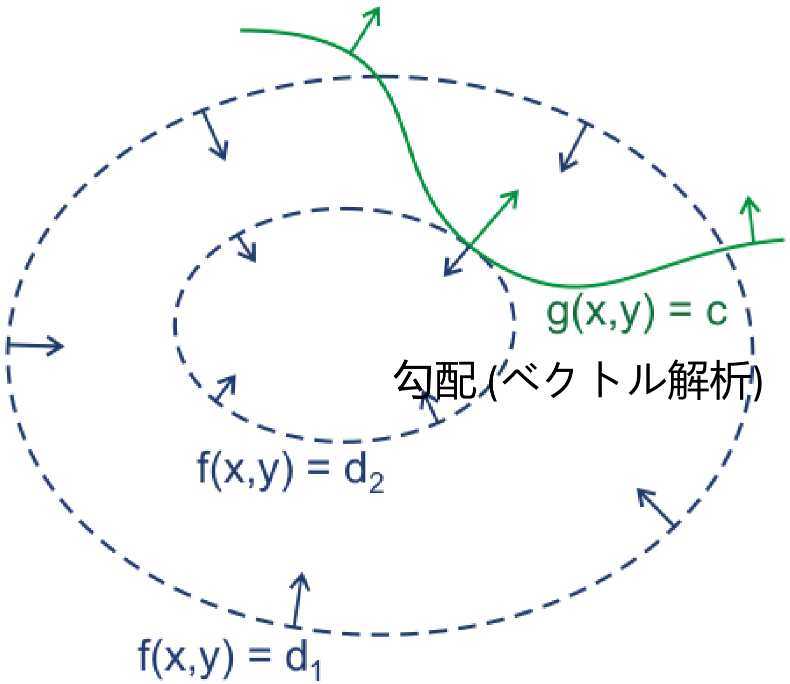


Lagrange multiplier

$$\min f(x,y)$$
$$s.t.\, g(x,y) = c$$

$$\phi(x,y,\lambda) = f(x,y) + \lambda g(x,y) = 0$$

$$\frac{\partial \phi}{\partial x} = 0; \frac{\partial \phi}{\partial y} = 0; \frac{\partial \phi}{\partial \lambda} = 0$$



Calculating this equations and discuss
the possible solution from its results.

derivation

Lagrange multiplier E.g.

$$\min f(x,y) = x^2 y$$

$$s.t. g(x,y) = x^2 + y^2 = 1$$

Please solve it!!!

Duality Theorem

- The problem and its Lagrange multiplier

$$\begin{aligned} &\min f(x) \\ &s.t. g(x) = 0 \end{aligned} \quad \Longrightarrow \quad \phi(x, \lambda) = f(x) + \lambda g(x) = 0$$

- Its dual problem
    - Advantage: change constraint problem to un- constraint problem

$$\max W(\lambda) = \phi(x(\lambda), \lambda)$$

$$x(\lambda) = \arg\min_{x} \phi(x, \lambda)$$

derivation

Duality Theorem E.g.

$$\min f(x, y) = x^2 + y^2$$

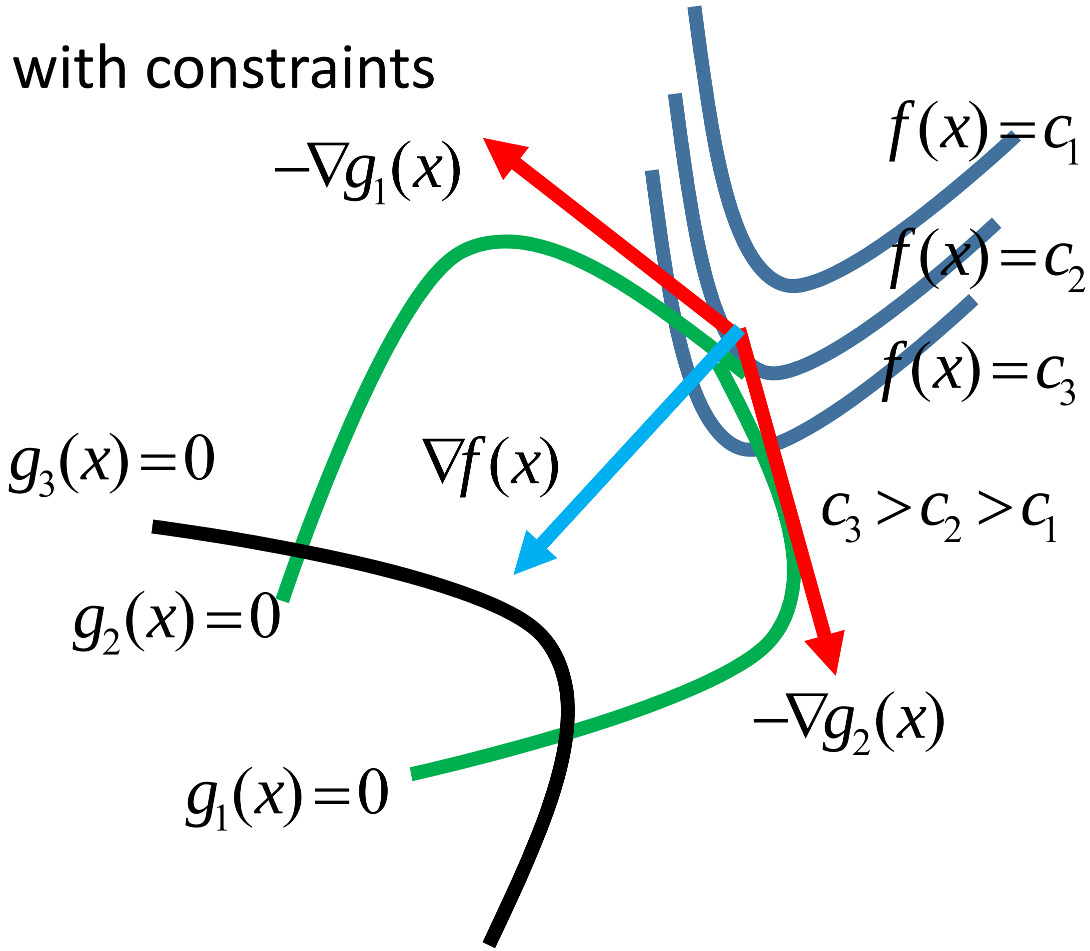
$$s.t. g(x, y): \quad x + y = 1$$

Please solve it by Duality Theorem!!!

Karush-Kuhn-Tucker Theorem (KKT)

- Given an optimization problem with constraints
  - Active/inactive constraint

$$\min f(x)$$
$$s.t.$$
$$g_1(x) \le 0,$$
$$g_2(x) \le 0,$$
$$\ldots,$$
$$g_n(x) \le 0$$



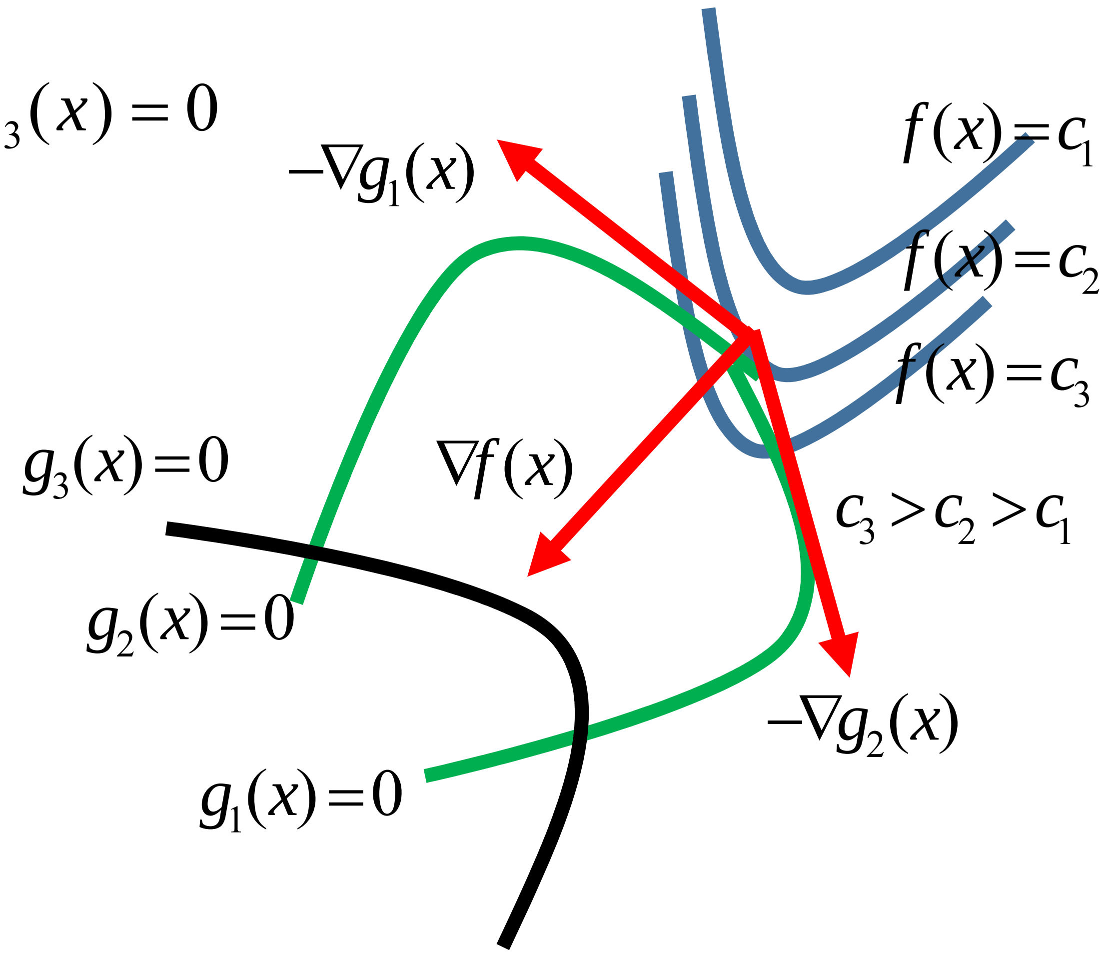
Karush-Kuhn-Tucker Theorem (KKT)

$$\nabla f(x) + \alpha_1 \nabla g_1(x) + \alpha_2 \nabla g_2(x) + \alpha_3 \nabla g_3(x) = 0$$

$$\alpha_1 > 0$$

$$\alpha_2 > 0$$

$$\alpha_3 = 0$$



Karush-Kuhn-Tucker Theorem (KKT)

$Let: f, g, h \in C$

$Let: x^{*}\_be\_a\_regular\_point$

$and\_a\_local\_minimizer\_for$

$problem\_f\_subject\_to$

$h(x) = 0, g(x) \le 0$

$Then, there\_exist\_\lambda^{*} \in R^{m} and\, \alpha^{*} \in R^{p}$

$(1)\alpha^{*} \ge 0$

$(2)\nabla f(x^{*}) + \lambda^{*T}\nabla h(x^{*}) + \alpha^{*T}\nabla g(x^{*}) = 0^{T}$

$(3)\alpha^{*T} g(x^{*}) = 0$

$(3.1)\alpha^{*} > 0 \rightarrow g(x^{*}) = 0$

$(3.2)g(x^{*}) < 0 \rightarrow \alpha^{*} = 0$

Karush-Kuhn-Tucker Theorem (KKT) E.g.

- Solve this problem with KKT

$$
\begin{aligned}
&\min \\
&x_1^2 - 6x_1 + 9 + x_2^2 + 2x_2 + 1 \\
&s.t. \\
&x_1 - x_2 \geq 1 \\
&x_1 + x_2 \leq 2
\end{aligned}
$$

Principal Component Analysis

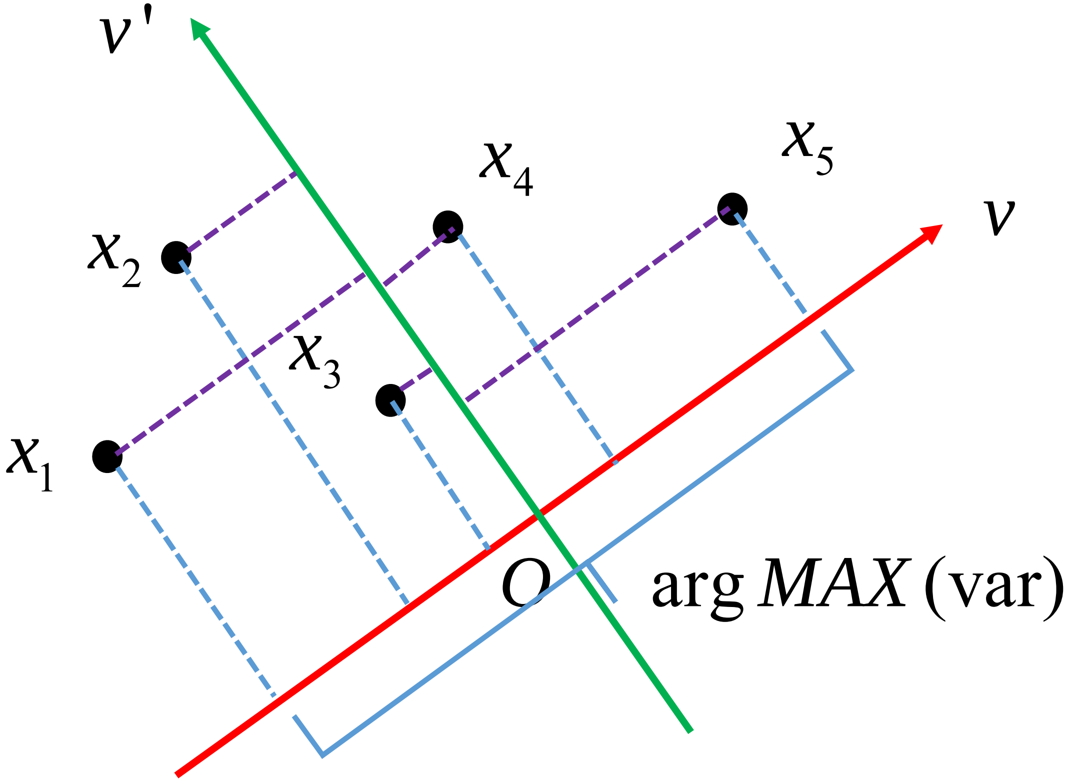
Principle Component Analysis

- Suppose there are training samples with zero mean:

$$x_1, x_2, \ldots, x_n \in R^d$$

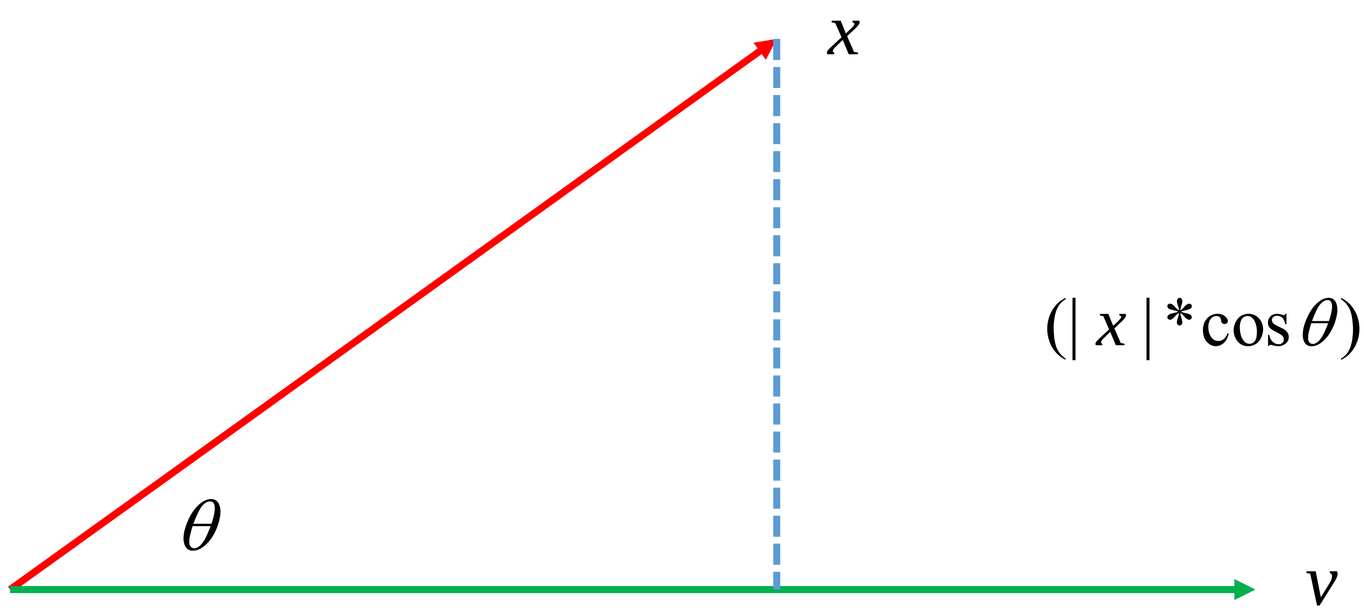
- Objective of PCA is to find a set of p<=d vectors in space R^d containing the maximum amount of variance in the data.

Principal Component Analysis



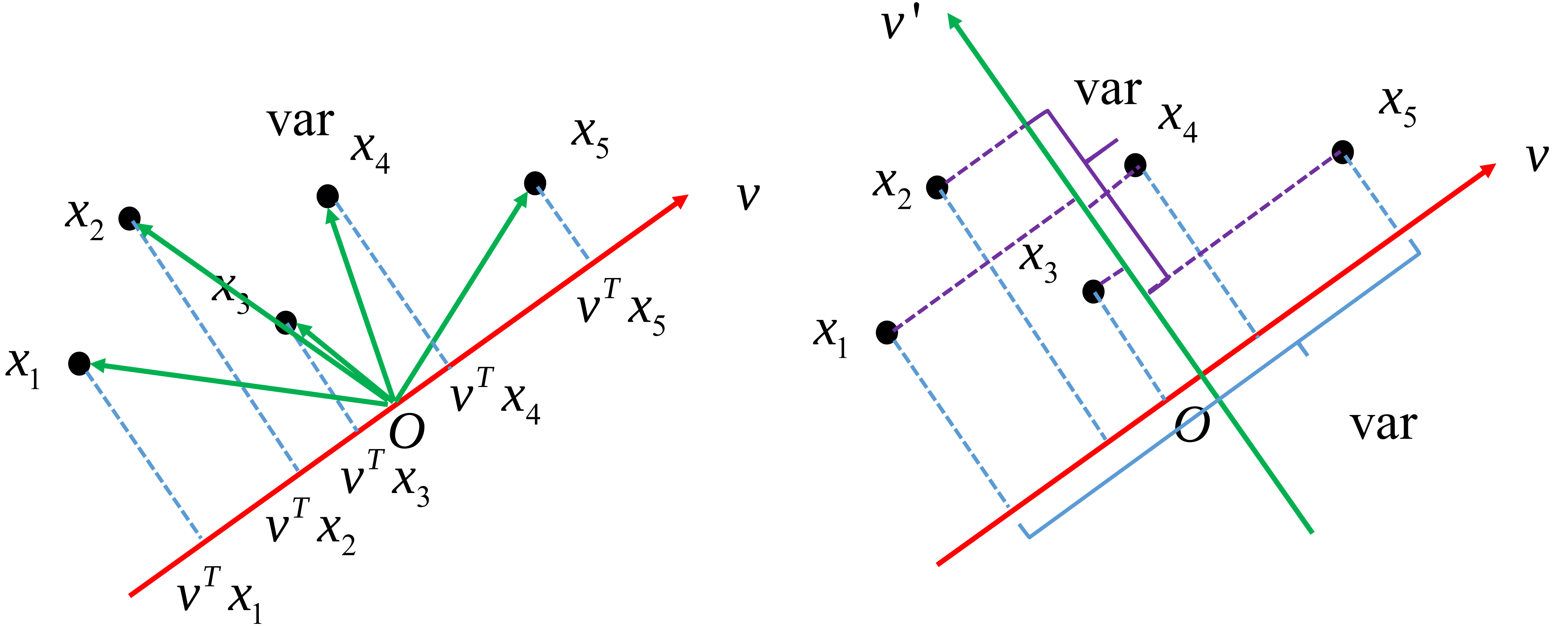
derivation

Principal Component Analysis



$$(|x|*\cos\theta)\frac{v}{|v|} = \ldots = \frac{<x,v>}{|v|^2}v$$

Principal Component Analysis



derivation

Principal Component Analysis

- The objective PCA is
  - Given projected vectors

$$v^T x_1, v^T x_2, \ldots, v^T x_N$$

  - Variance

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(v^T x_{\mathrm{i}} \text{-} 0)^2 = \ldots = v^T [\frac{1}{N}\sum_{i=1}^{N} x_{\mathrm{i}} x_i^T] v$$

$$C = \frac{1}{N}\sum_{i=1}^{N} x_{\mathrm{i}} x_i^T$$

  - First principal vector is

$$v = \arg \max_{v \in R^d, |v|=1} \mathrm{v}^T C \mathrm{v}$$

derivation

Principal Component Analysis

- How to solve $v = \arg\max_{v \in R^d, |v|=1} v^T C v$ C: matrix

$$f(v, \lambda) = v^T C v - \lambda (v^T v - 1)$$

- Lagrangian

$$\frac{\partial f}{\partial v} = 0 \rightarrow$$

$$\frac{\partial f}{\partial \lambda} = 0 \rightarrow$$

Principal Component Analysis

- Note: covariance matrix

$$C = \frac{1}{N}\sum_{i=1}^{N} x_i\, x_i^T = \frac{1}{N}[x_1, \ldots, x_n]\begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix}$$

$$C = \frac{1}{N} X^T X$$