Statistique et probabilités

1. Motivation : estimation problems
2. Computing probabilities
3. Computing and controlling moments and tail probabilities

Moments

Variance $\mathrm{Var}[X] = E[(X - E[X])^2]$

Moment $E[X^k]$

Moment centré $E[(X - E[X])^k]$

Covariance $\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$

Linéarité de l'espérance

$$E\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i E(X_i)$$

Bilinéarité de la covariance

$$\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \mathrm{Cov}(X_i, Y_j)$$

Moments conditionnels

Espérance conditionnelle

$$f(y) = E[X|Y=y] = \int x P_{X|Y=y}(dx)$$

Variance conditionnelle

$$g(y) = \mathrm{Var}[X|Y=y] = E[(X-E[X])^2|Y=y]$$

Loi de l'espérance totale

$$E[X] = E[E[X|Y]]$$

Loi de la covariance totale

$$\mathrm{Cov}[X,Y] = \mathrm{Cov}[E[X|Z], E[Y|Z]] + E[\mathrm{Cov}[X,Y|Z]]$$

Statistique et probabilités

1. Motivation : estimation problems
2. Computing probabilities
3. Computing and controlling moments and tail probabilities

Algèbre linéaire

Algèbre linéaire

1. Espaces vectoriels et fonctions linéaires
2. Matrices
3. Angles et orthogonalité
4. Structure des applications linéaires

Algèbre linéaire

1. Espaces vectoriels et fonctions linéaires
2. Matrices
3. Angles et orthogonalité
4. Structure des applications linéaires

Fonction linéaire

- Objet central de l'algèbre linéaire (du point de vue conceptuel)
- Idée générale: fonction qui préserve les combinaisons linéaires
- Définition formelle

$$f : \left| \begin{array}{ccc} E & \to & F \\ x & \mapsto & f(x) \end{array} \right.$$ est une fonction linéaire si et seulement si

E et F sont des espaces vectoriels sur un même corps K et pour toute paire de scalaires $(\alpha,\, \beta) \in K^2$ et toute paire de vecteurs $(u,\, v) \in E^2$,

$$f(\alpha u + \beta v) = \alpha f(u) + \beta f(v)$$

Décomposition sur une base

Définition (décomposition sur une base)

Soit une base $V = (v_i)_{i \in I}$ de E et $\Lambda = (\lambda_i)_{i \in I} \in K^I$ une famille de scalaires telle que au plus un nombre fini des λ_i soient différents de 0.

On dit que Λ est une décomposition de $v \in E$ sur la base V si et seulement si $v = \sum_{i \in I} \lambda_i v_i$.

Théorème (existence et unicité de la décomposition)

Etant donné une base $V = (v_i)_{i \in I}$ de E, tout vecteur $v \in E$ peut-être décomposé sur cette base et cette décomposition est unique.

Toute base peut donc servir de système de coordonnées pour E

Algèbre linéaire

1. Espaces vectoriels et fonctions linéaires
2. Matrices
3. Angles et orthogonalité
4. Structure des applications linéaires

Matrice

- Objet central de l'algèbre linéaire (en pratique)
- Idée générale: simplement un tableau de nombres en deux dimensions

$$\mathrm{M} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{np} \end{pmatrix}$$

Matrice

- Objet central de l'algèbre linéaire (en pratique)
- Idée générale: simplement un tableau de nombres en deux dimensions
- Pourquoi est-ce utile ?

$$M = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{np} \end{pmatrix}$$

Matrice

- Objet central de l'algèbre linéaire (en pratique)
- Idée générale: simplement un tableau de nombres en deux dimensions
- Pourquoi est-ce utile ?
  - Permet de représenter les fonctions linéaires en dimensions finie de manière simple à appréhender et pratique pour les calculs

$$\mathrm{M} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{np} \end{pmatrix}$$

Matrice d’une fonction linéaire

E, F espaces vectoriels de dimension finie
$V = (v_1, ..., v_p)$ base de E
$W = (w_1, ..., w_n)$ base de F
$f : E \to F$, fonction linéaire
$M = M(f, V, W)$ matrice de f de V vers W.

$$\mathrm{M} = \begin{array}{cc}
\begin{pmatrix}
a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1p} \\
a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2p} \\
\vdots & \vdots & & \vdots & & \vdots \\
a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{ip} \\
\vdots & \vdots & & \vdots & & \vdots \\
a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{np}
\end{pmatrix} &
\begin{matrix} \leftarrow w_1 \\ \leftarrow w_2 \\ \\ \leftarrow w_i \\ \\ \leftarrow w_n \end{matrix} \\
\begin{matrix} \uparrow & \uparrow & & \uparrow & & \uparrow \\ f(v_1) & f(v_2) & & f(v_j) & & f(v_p) \end{matrix} &
\end{array}$$

Matrice d'une fonction linéaire

Multiplication matrice/vecteur colonne: $(Mv)_i := \sum_{j=1}^{n} m_{ij} v_i$

Multiplication matrice-matrice: $(AB)_{ij} := \sum_{k=1}^{n} a_{ik} b_{kj}$

Matrice d’une fonction linéaire

L’évaluation en un point d’une fonction linéaire et la composition d’application linéaire se réduit (en dimension finie) à l’application “mécanique” de règles de calcul matriciel

Pas trop “mécanique” quand même:

Crimes contre les matrices
http://ee263.stanford.edu/notes/matrix_crimes.pdf

Changement de base

Toute base peut servir de système de coordonnées pour E

Changement de base == changement de système de coordonnées

Matrice de passage de V à W: $P(V, W) = M(Id_E, V, W)$

Définition: transforme un vecteur v exprimé en terme de ses coordonnées dans V, en le même vecteur exprimé en terme de ses coordonnées dans W

Pour la trouver: les colonnes de P(V,W) sont les coordonnées des éléments de V (dans un ordre fixé) dans W

Matrice par blocs

La matrice

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & 2 & 2 & 2 \\ 1 & 1 & 2 & 2 & 2 \\ 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 4 & 4 & 4 \end{bmatrix}$$

peut être partitionnée en quatre blocs

$$\mathbf{P}_{11} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \mathbf{P}_{12} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}, \mathbf{P}_{21} = \begin{bmatrix} 3 & 3 \\ 3 & 3 \\ 3 & 3 \end{bmatrix}, \mathbf{P}_{22} = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

On peut alors écrire la matrice par bloc comme :

$$\mathbf{P}_{\text{partitionnee}} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}.$$

Interpretation des opérations matricielles

write $A \in \mathbf{R}^{m \times n}$ in terms of its columns:

$$A = \left[\begin{array}{cccc} a_1 & a_2 & \cdots & a_n \end{array} \right]$$

then $y = Ax$ can be written as

$$y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

write A in terms of its rows:

$$A = \left[\begin{array}{c} \tilde{a}_1^T \\ \tilde{a}_2^T \\ \vdots \\ \tilde{a}_n^T \end{array} \right]$$

then $y = Ax$ can be written as

$$y = \left[\begin{array}{c} \tilde{a}_1^T x \\ \tilde{a}_2^T x \\ \vdots \\ \tilde{a}_m^T x \end{array} \right]$$

Interpretation des opérations matricielles

$$c_{ij} = \tilde{a}_i^T b_j = \langle \tilde{a}_i, b_j \rangle$$

$$C = \left[\begin{array}{c} c_1 \cdots c_p \end{array}\right] = AB = \left[\begin{array}{c} Ab_1 \cdots Ab_p \end{array}\right]$$

$$C = \left[\begin{array}{c} \tilde{c}_1^T \\ \vdots \\ \tilde{c}_m^T \end{array}\right] = AB = \left[\begin{array}{c} \tilde{a}_1^T B \\ \vdots \\ \tilde{a}_m^T B \end{array}\right]$$

$$C = \sum_i a_i \tilde{b}_i^T$$

Algèbre linéaire

1. Espaces vectoriels et fonctions linéaires

2. Matrices

3. Angles et orthogonalité

4. Structure des applications linéaires et matrices

Produit scalaire usuel

$$(a \mid b) = \sum_{i=1}^{n} a_i bi = a^T b = \|a\|_2 \|b\|_2 \cos(\theta)$$

Angles et orthogonalité

Intérêt ?

Soit E un espace vectoriel de dimension finie n muni d'un produit scalaire $(\,.\mid.\,)$ et $V = (v_1, \ldots, v_n)$ une base orthonormale de E. Alors, pour tout $v \in E$,

$$v = \sum_{i=1}^{n} (v \mid v_i) v_i.$$

Algèbre linéaire

1. Espaces vectoriels et fonctions linéaires

2. Matrices

3. Angles et orthogonalité

4. Structure des applications linéaires et matrices

Décomposition en valeurs singulières

Etant donné une matrice rectangulaire quelconque

(à coefficients réels)

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}$$

Quelle transformation linéaire représente-t-elle ?

Par example : $M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$

Décomposition en valeurs singulières

$$\boldsymbol{AV = U\Sigma} \qquad A\begin{bmatrix} \\ \boldsymbol{v}_1 \,..\, \boldsymbol{v}_r \,..\, \boldsymbol{v}_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{u}_1 \,..\, \boldsymbol{u}_r \,..\, \boldsymbol{u}_m \\ \\ \end{bmatrix} \left[\begin{array}{ccc|c} \boldsymbol{\sigma_1} & & & \\ & \ddots & & \mathbf{0} \\ & & \boldsymbol{\sigma_r} & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right]$$

Flashback: interpretation des opérations matricielles

$$c_{ij} = \tilde{a}_i^T b_j = \langle \tilde{a}_i, b_j \rangle$$

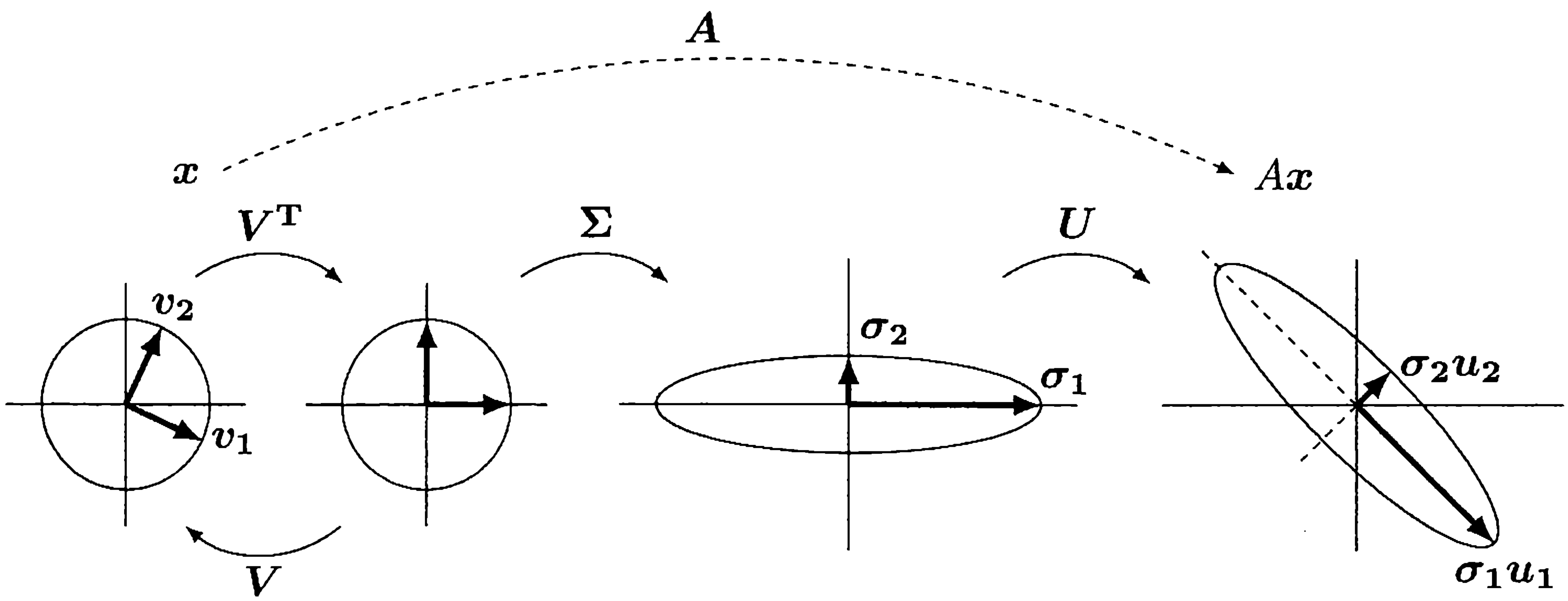
$$C = \left[\begin{array}{c} c_1 \cdots c_p \end{array}\right] = AB = \left[\begin{array}{c} Ab_1 \cdots Ab_p \end{array}\right]$$

$$C = \left[\begin{array}{c} \tilde{c}_1^T \\ \vdots \\ \tilde{c}_m^T \end{array}\right] = AB = \left[\begin{array}{c} \tilde{a}_1^T B \\ \vdots \\ \tilde{a}_m^T B \end{array}\right]$$

$$C = \sum_i a_i \tilde{b}_i^T$$

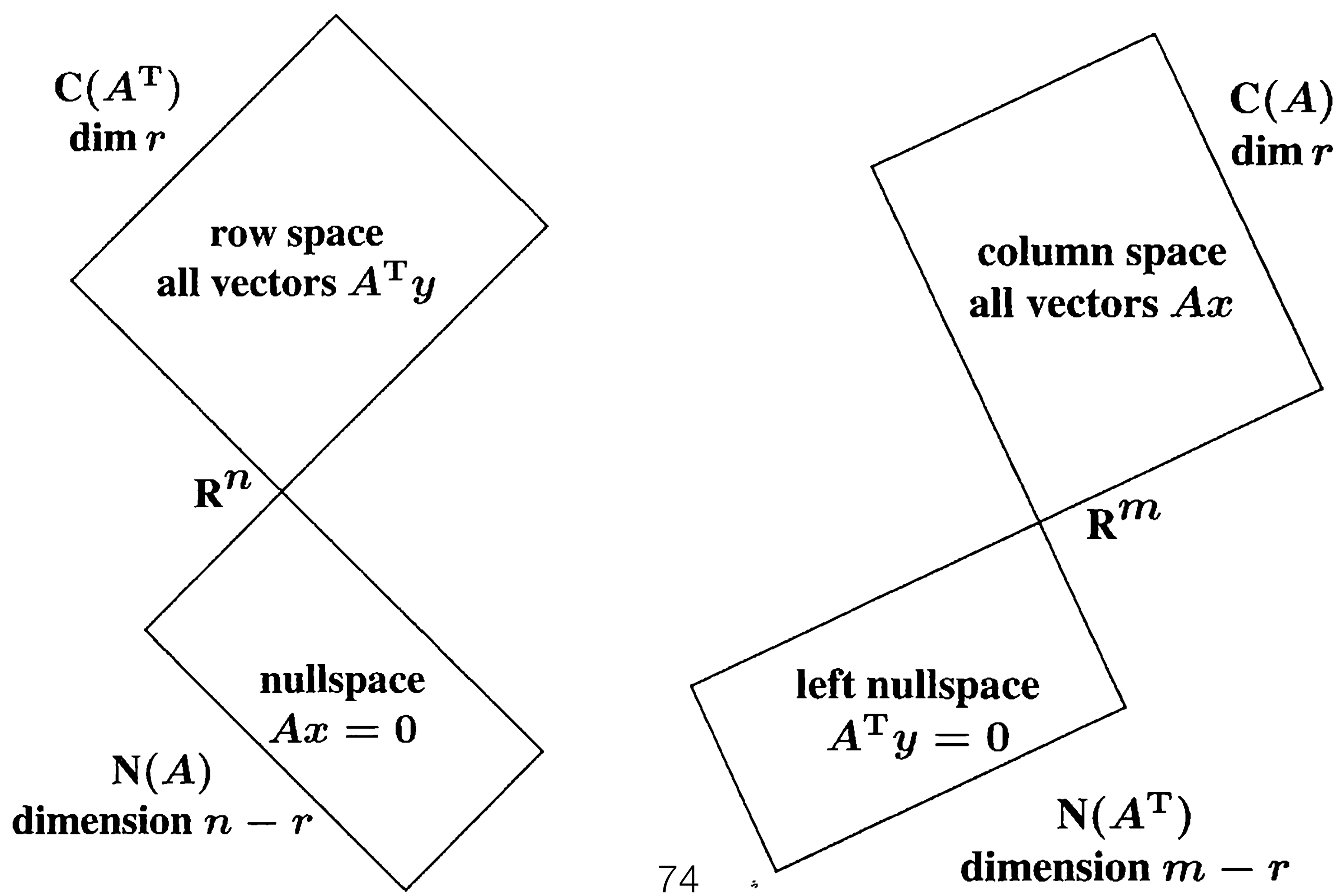
Décomposition en valeurs singulières

$$
\boxed{AV = U\Sigma \qquad A\begin{bmatrix} \\ v_1 \, .. \, v_r \, .. \, v_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ u_1 \, .. \, u_r \, .. \, u_m \\ \\ \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right]}
$$



Espaces associés à une matrice

$$\boldsymbol{AV = U\Sigma} \qquad A\begin{bmatrix} \\ \boldsymbol{v}_1 \;..\; \boldsymbol{v}_r \;..\; \boldsymbol{v}_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{u}_1 \;..\; \boldsymbol{u}_r \;..\; \boldsymbol{u}_m \\ \\ \end{bmatrix} \left[\begin{array}{ccc|c} \boldsymbol{\sigma_1} & & & \\ & \ddots & & \mathbf{0} \\ & & \boldsymbol{\sigma_r} & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right]$$



Espaces associés à une matrice

$$\boldsymbol{AV = U\Sigma} \qquad A\begin{bmatrix} \\ \boldsymbol{v}_1 \, .. \, \boldsymbol{v}_r \, .. \, \boldsymbol{v}_n \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{u}_1 \, .. \, \boldsymbol{u}_r \, .. \, \boldsymbol{u}_m \\ \\ \end{bmatrix} \left[\begin{array}{ccc|c} \boldsymbol{\sigma_1} & & & \\ & \ddots & & \mathbf{0} \\ & & \boldsymbol{\sigma_r} & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right]$$

Par exemple : $M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$

$$U \approx \begin{pmatrix} -.21 & -.89 & .41 \\ -.52 & -.25 & -.82 \\ -.83 & .39 & .41 \end{pmatrix} \qquad V \approx \begin{pmatrix} -.48 & -.57 & -.66 \\ .78 & .08 & -.62 \\ .41 & -.82 & .41 \end{pmatrix}$$

$$\begin{pmatrix} \sigma_1 & \sigma_2 & \sigma_3 \end{pmatrix} \approx \begin{pmatrix} 16.85 & 1.07 & 0 \end{pmatrix}$$

Théorème spectral

Si S est une matrice symétrique, réelle de taille $m \times m$, alors il existe une matrice orthogonale réelle Q de taille $m \times m$ et une matrice diagonale réelle Λ de taille $m \times m$ telles que $S = Q\Lambda Q^T$.

Positivité

Une matrice symétrique réelle est dite définie positive, noté $S \succ 0$ ssi pour toute matrice colonne u, $u^T S u > 0$.

Une matrice symétrique réelle est dite semi-définie positive, $S \succeq 0$, ssi pour toute matrice colonne u, $u^T S u \geq 0$.

Relation d'ordre sur les matrices symétriques réelles:
$S_1 \prec S_2$ ssi $S_2 - S_1 \succ 0$
et
$S_1 \preceq S_2$ ssi $S_2 - S_1 \succeq 0$

Déterminant et trace

A matrice carrée $n \times n$

$$\det(A) = \sum_{\sigma \in S_n} \left(\operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma_i} \right),$$

sgn(σ) est la parité du nombre d'élement dans une décomposition de σ en une séquence de transpositions (échange de deux éléments).

$$\det(AB) = \det(A)\det(B)$$

$$\operatorname{Tr}(A) = \sum_{i=1}^{n} a_{i,i} \quad \operatorname{Tr}(\mathrm{AB}) = \operatorname{Tr}(\mathrm{BA})$$

$$\operatorname{Tr}(A_1 A_2 \ldots A_k) = \operatorname{Tr}(A_2 A_3 \ldots A_k A_1) = \cdots = \operatorname{Tr}(A_k A_1 A_2 \ldots A_{k-1})$$

Algèbre linéaire

1. Espaces vectoriels et fonctions linéaires
2. Matrices
3. Angles et orthogonalité
4. Structure des applications linéaires et matrices

Optimisation stochastique

Contexte : minimisation du risque empirique pour une fonction de coût "séparable par point de donnée"

$$R_n(w) = \frac{1}{n}\sum_{i=1}^{n} f_i(w)$$

Descente de gradient stochastique

$$w_1 \in \mathbb{R}^d \text{ given}$$

$$w_{k+1} \leftarrow w_k - \alpha_k \nabla f_{i_k}(w_k)$$

i_k is chosen *randomly* from $\{1, \ldots, n\}$ and α_k is a positive stepsize

Optimisation stochastique

Exemple de garantie de convergence

(cf. Bottou, Curtis et Nocedal (2018) Optimisation Methods for Large-Scale Machine Learning)

Si

$$\sum_{k=1}^{\infty} \alpha_k = \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \alpha_k^2 < \infty$$

Assumption 4.1 (Lipschitz-continuous objective gradients). *The objective function* $F : \mathbb{R}^d \to \mathbb{R}$ *is continuously differentiable and the gradient function of* F, *namely,* $\nabla F : \mathbb{R}^d \to \mathbb{R}^d$, *is Lipschitz continuous with Lipschitz constant* $L > 0$, *i.e.,*

$$\|\nabla F(w) - \nabla F(\overline{w})\|_2 \leq L\|w - \overline{w}\|_2 \;\; \text{for all} \;\; \{w, \overline{w}\} \subset \mathbb{R}^d.$$

plus des conditions de régularité pas très contraignantes

Alors

$$\liminf_{k \to \infty} \mathbb{E}[\|\nabla F(w_k)\|_2^2] = 0$$

Example Distributions

Distribution	PDF or PMF	Mean	Variance
$\mathit{Bernoulli}(p)$	$\begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{if } x = 0. \end{cases}$	p	$p(1-p)$
$\mathit{Binomial}(n,p)$	$\binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, ..., n$	np	$np(1-p)$
$\mathit{Geometric}(p)$	$p(1-p)^{k-1}$ for $k = 1, 2, ...$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$\mathit{Poisson}(\lambda)$	$\frac{e^{-\lambda}\lambda^k}{k!}$ for $k = 0, 1, ...$	λ	λ
$\mathit{Uniform}(a,b)$	$\frac{1}{b-a}$ for all $x \in (a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathit{Gaussian}(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ for all $x \in (-\infty, \infty)$	μ	σ^2
$\mathit{Exponential}(\lambda)$	$\lambda e^{-\lambda x}$ for all $x \geq 0, \lambda \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

[2] Table reproduced from Maleki & Do's review handout by Koochak & Irvin

Random Vectors

Given n RV's $X_1, ..., X_n$, we can define a random vector X s.t.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

Note: all the notions of joint PDF/CDF will apply to X.

Given $g : \mathbb{R}^n \to \mathbb{R}^m$, we have:

$$g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_m(x) \end{bmatrix}, \mathbb{E}[g(X)] = \begin{bmatrix} \mathbb{E}[g_1(X)] \\ \mathbb{E}[g_2(X)] \\ \vdots \\ \mathbb{E}[g_m(X)] \end{bmatrix}.$$

Covariance Matrices

For a random vector $X \in \mathbb{R}^n$, we define its **covariance matrix** Σ as the $n \times n$ matrix whose ij-th entry contains the covariance between X_i and X_j.

$$\Sigma = \begin{bmatrix} Cov[X_1, X_1] & \dots & Cov[X_1, X_n] \\ \vdots & \ddots & \vdots \\ Cov[X_n, X_1] & \dots & Cov[X_n, X_n] \end{bmatrix}$$

applying linearity of expectation and the fact that $Cov[X_i, X_j] = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$, we obtain

$$\Sigma = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]$$

Properties:

- Σ is symmetric and PSD
- If $X_i \perp X_j$ for all i, j, then $\Sigma = diag(Var[X_1], ..., Var[X_n])$

Multivariate Gaussian

The multivariate Gaussian $X \sim \mathcal{N}(\mu, \Sigma)$, $X \in \mathbb{R}^n$:

$$p(x; \mu, \Sigma) = \frac{1}{det(\Sigma)^{\frac{1}{2}}(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

The univariate Gaussian $X \sim \mathcal{N}(\mu, \sigma^2)$, $X \in \mathbb{R}$ is just the special case of the multivariate Gaussian when $n = 1$.

$$p(x; \mu, \sigma^2) = \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

Notice that if $\Sigma \in \mathbb{R}^{1\times 1}$, then $\Sigma = Var[X_1] = \sigma^2$, and so

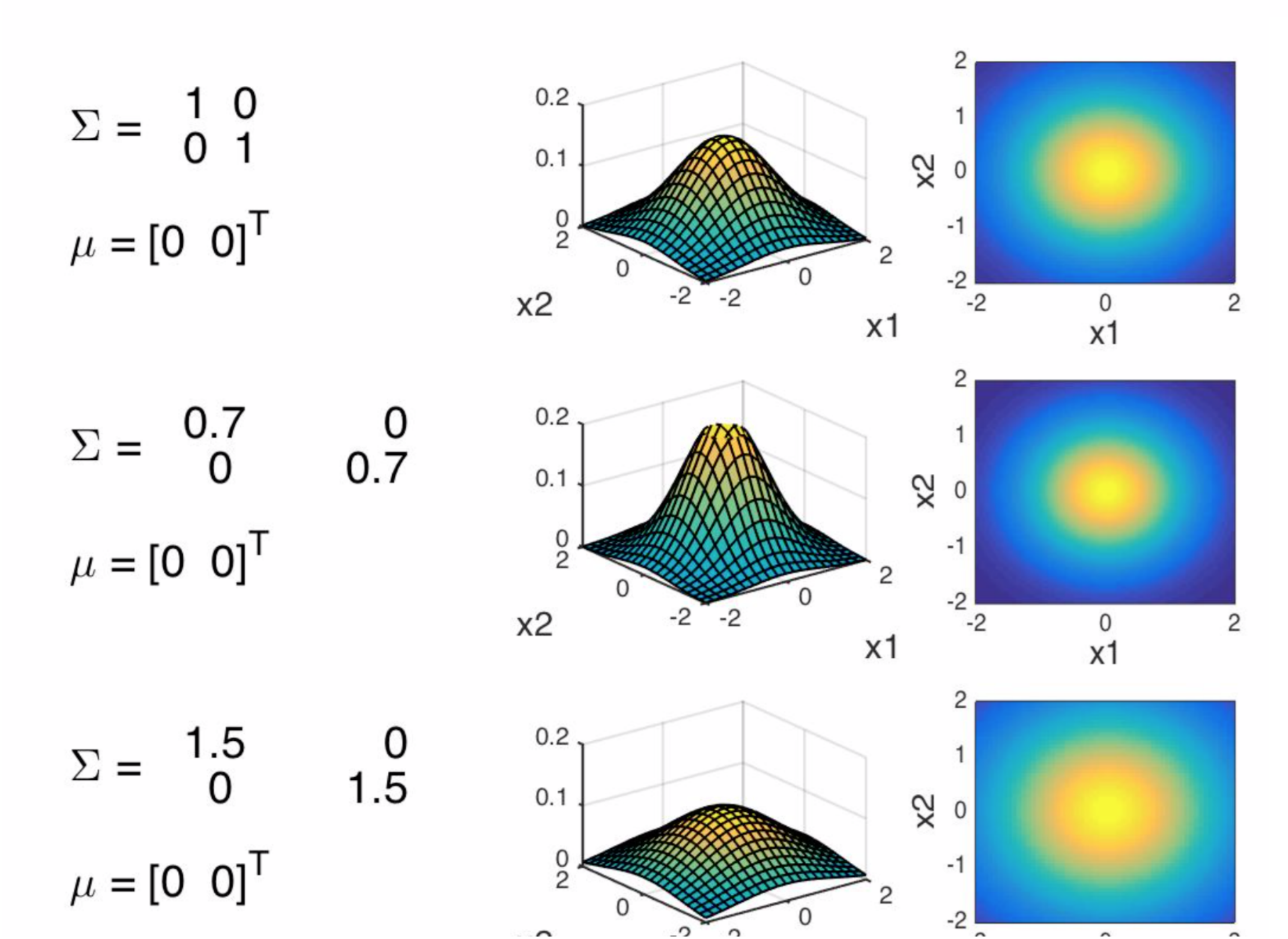
- $\Sigma^{-1} = \frac{1}{\sigma^2}$
- $det(\Sigma)^{\frac{1}{2}} = \sigma$

Some Nice Properties of MV Gaussians

- Marginals and conditionals of a joint Gaussian are Gaussian
- A d-dimensional Gaussian $X \in \mathcal{N}(\mu, \Sigma = diag(\sigma_1^2, ..., \sigma_n^2))$ is equivalent to a collection of d **independent** Gaussians $X_i \in \mathcal{N}(\mu_i, \sigma_i^2)$. This results in isocontours aligned with the coordinate axes.
- In general, the isocontours of a MV Gaussian are n-dimensional ellipsoids with principal axes in the directions of the eigenvectors of covariance matrix Σ (remember, Σ is PSD, so all n eigenvectors are non-negative). The axes' relative lengths depend on the eigenvalues of Σ.

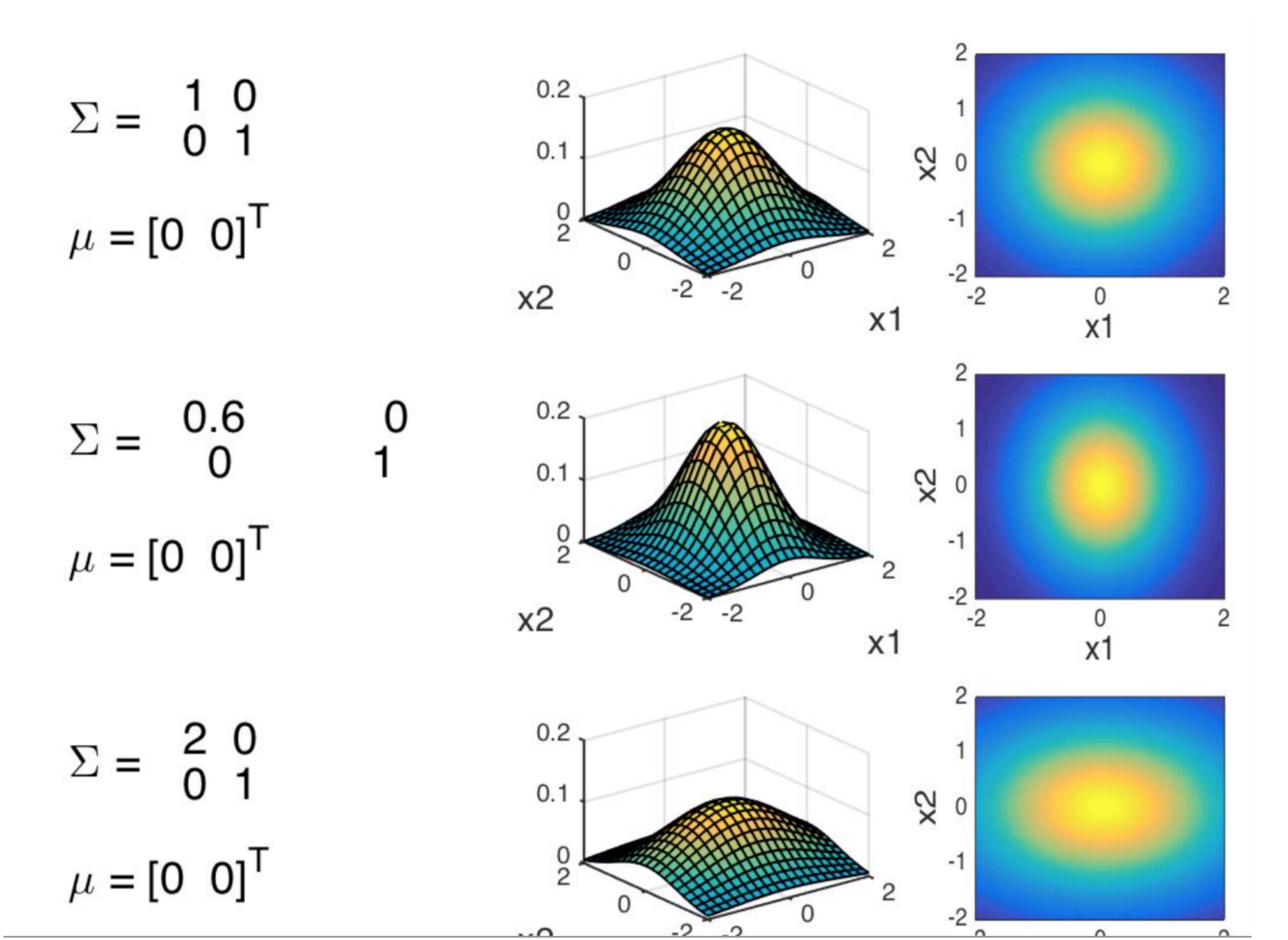
Visualizations of MV Gaussians

Effect of changing variance



Visualizations of MV Gaussians

If $Var[X_1] \neq Var[X_2]$:



Visualizations of MV Gaussians

If X_1 and X_2 are positively correlated:

$$\Sigma = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

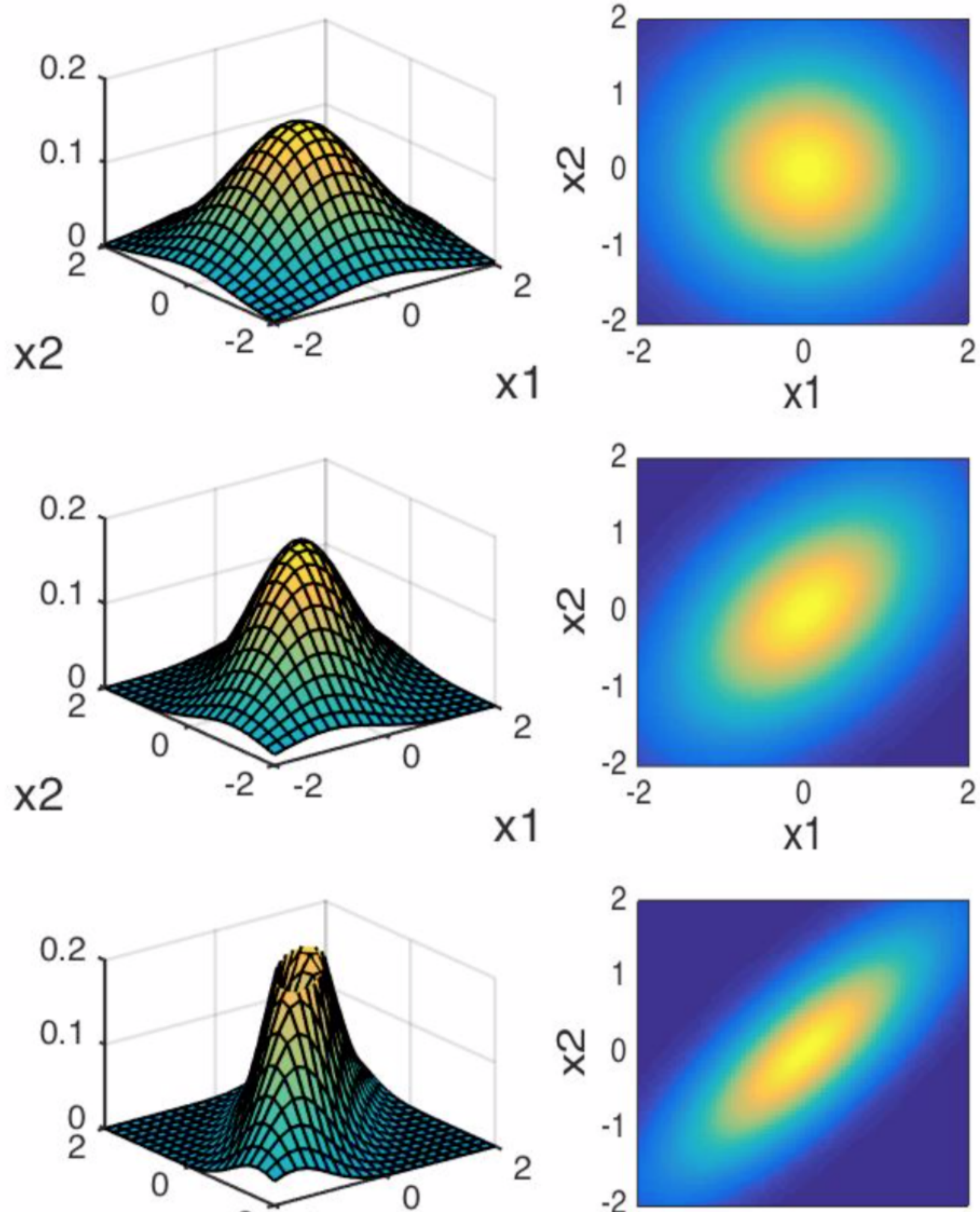
$$\mu = [0 \;\; 0]^T$$

$$\Sigma = \begin{matrix} 1 & 0.5 \\ 0.5 & 1 \end{matrix}$$

$$\mu = [0 \;\; 0]^T$$

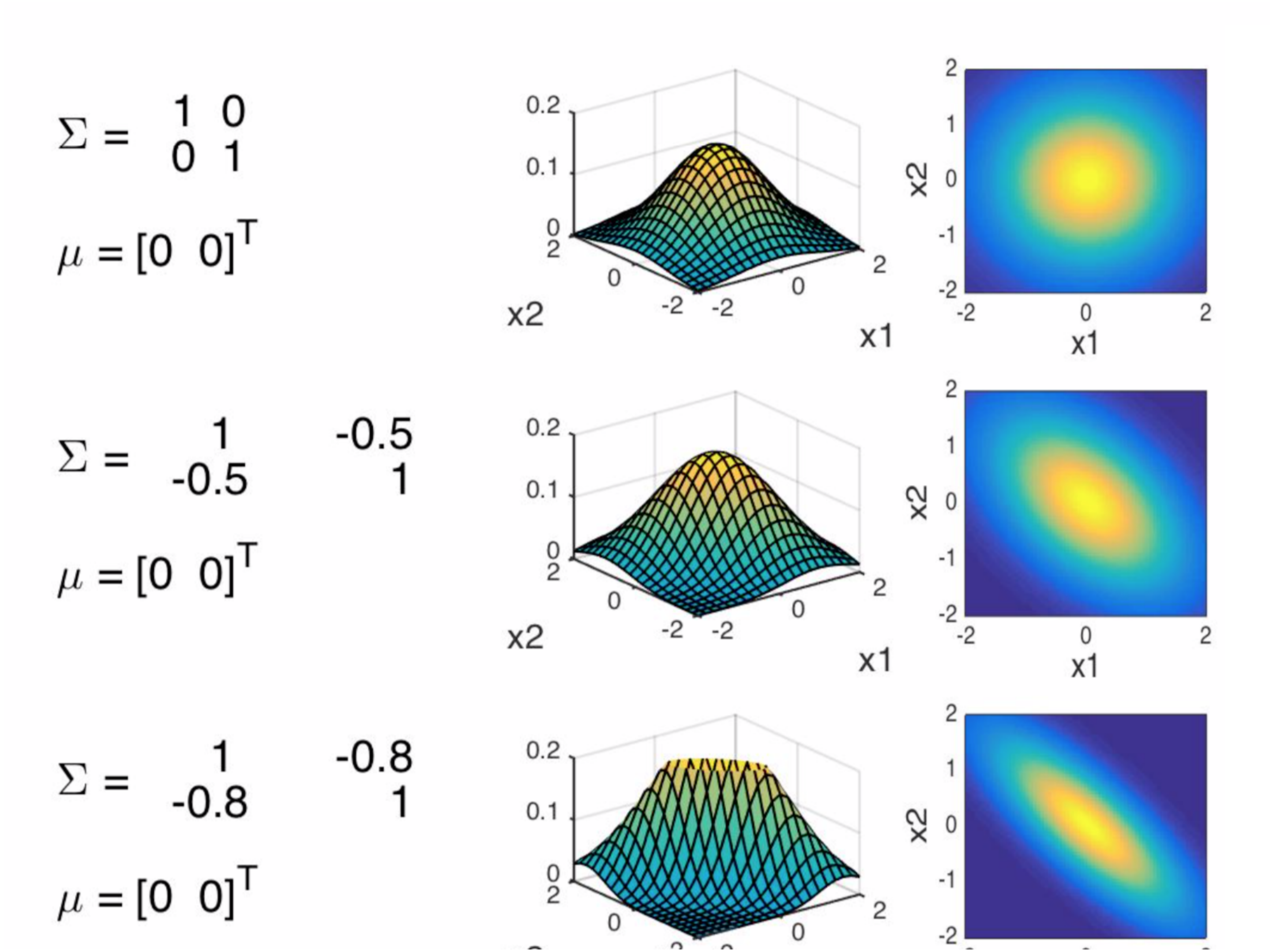
$$\Sigma = \begin{matrix} 1 & 0.8 \\ 0.8 & 1 \end{matrix}$$

$$\mu = [0 \;\; 0]^T$$



Visualizations of MV Gaussians

If X_1 and X_2 are negatively correlated:



Multivariate Gaussian

Définition générale

$$\mathbf{X} \sim \mathcal{N}(\mu, \mathbf{\Sigma}) \iff \text{there exist } \mu \in \mathbb{R}^k, \boldsymbol{A} \in \mathbb{R}^{k\times\ell} \text{ such that } \mathbf{X} = \boldsymbol{A}\mathbf{Z} + \mu \text{ for } Z_n \sim \mathcal{N}(0,1), \text{i.i.d.}$$

Distributions conditionnelles

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times 1 \\ (N-q) \times 1 \end{bmatrix} \qquad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times 1 \\ (N-q) \times 1 \end{bmatrix}$$

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times q & q \times (N-q) \\ (N-q) \times q & (N-q) \times (N-q) \end{bmatrix}$$

$$p(\mathbf{x}_1 \mid \mathbf{x}_2 = \mathbf{a}) = \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\mathbf{\Sigma}}), with$$

$$\bar{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{+}(\mathbf{a} - \boldsymbol{\mu}_2)$$

$$\bar{\mathbf{\Sigma}} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{+}\mathbf{\Sigma}_{21}$$

Distributions marginales ?