Gaussian Random Vectors

- very important in detection / estimation / stoch. processes
  1) Noise reasonably modeled as Gaussian
  2) easy to ~~work~~ work with – involves only means and covariances
  3) Gaussian case bounds performance for other r.v.'s with same means / covariances.

     MMSE estimator for Gauss. case same and has same MSE as LLSE estimator for other probs with same mean and covariance.

     → simple and linear in observations.

     MMSE estimator for non-Gaussian case has better performance than that for Gauss. problems with same mean / covariance.

     Gauss. case illuminates other probs.

Gaussian r.v.

W is normalized Gaussian r.v. if pdf

$$f_W(w) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2}$$

Easy to verify $E[W] = 0$, $\mathrm{var}[W] = 1$.

If $Z = \sigma W$, then $\underset{F_Z(z)}{\underset{\parallel}{P(Z \le z)}} = P(\sigma W \le z) = \underset{F_W(\frac{z}{\sigma})}{\underset{\parallel}{P(W \le \frac{z}{\sigma})}}$

$$\cancel{f_Z(z) = \frac{d}{} \quad f_W(w) = \frac{d}{dw} P(W \le w)}$$

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} F_W\left(\frac{z}{\sigma}\right) = \frac{1}{\sigma} f_W\left(\frac{z}{\sigma}\right)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{z^2}{2\sigma^2}}$$ 0-mean Gaussian r.v.

$\mathrm{var}(Z) = \sigma^2$ As $\sigma^2 \to 0$: density $\to \delta(z)$ impulse

i.e. $P(Z = 0) = 1$.

$z \xrightarrow{\text{shift}} u = z + m \qquad E[u] = m.$

$$f_u(u) = f_z(u-m) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(u-m)^2}{2\sigma^2}\right]$$

write : $u \sim \mathcal{N}(m, \sigma^2)$

- often easier to work with the 0-mean part : $u = m_u + \tilde{u}$
  where $\tilde{u}$ is 0-mean Gauss. r.v.
- Moment Generating fn. (MGF) of Gaussian r.v. $z \sim \mathcal{N}(0, \sigma^2)$

$$g_z(s) = E[e^{sz}] = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} e^{sz} e^{-\frac{z^2}{2\sigma^2}}\, dz$$

$$= e^{s^2\sigma^2/2}$$

by completing square in exponent (see notes)

- MGF used to generate moments of z
- can verify (Ex. 2.2)

$$E[z^{2k}] = \frac{(2k)!\,\sigma^{2k}}{k!\,2k} = (2k-1)(2k-3)(2k-5)\cdots(3)(1)\,\sigma^{2k}$$

$$E[z^4] = 3\sigma^4, \quad E[z^6] = 15\sigma^6 \ldots$$

- z^{2k+1} odd fn. and Gauss. is even. Thus,
  $E[z^{2k+1}] = 0$ all $k = 0, 1, \ldots$ odd moments $= 0$.
- For $u \sim \mathcal{N}(m, \sigma^2)$, $u = m + \underset{\substack{\uparrow \\ \text{zero-mean}}}{\tilde{u}}$

$$g_u(s) = E[e^{s(m+\tilde{u})}] = E[e^{sm} e^{s\tilde{u}}] = e^{sm} e^{s^2\sigma^2/2}.$$

- $f_u(u)$ uniquely determined by $g_u(s)$. (up to set of measure 0)
  Thus $f_u(u) = \mathcal{N}(m, \sigma^2)$ iff $g_u(s) = \exp\left(sm + \frac{s^2\sigma^2}{2}\right)$

Gaussian Random Vectors and MGF's

- $\underline{z} : \Omega \to \mathbb{R}^n$
- view as vector of r.v.'s $\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \underline{z}$

Prob. density $f_{\underline{z}}(\underline{z})$ – joint pdf of components $z_1, \ldots, z_n$.

Mean $E[\underline{z}] = \begin{bmatrix} m_{z_1} \\ \vdots \\ m_{z_n} \end{bmatrix}$, $m_{z_i} = E[z_i]$

$\qquad \| $

$\quad \underline{m}_z$

covariance $\underset{n \times n}{K_{\underline{z}}} = E[(\underline{z} - \underline{m}_z)(\underline{z} - \underline{m}_z)^T]$ $\quad (K_z)_{ij} = E[(z_i - m_{z_i})(z_j - m_{z_j})]$

MGF $\quad g_{\underline{z}}(\underline{s}) = E[\exp(\underline{s}^T \underline{z})]$

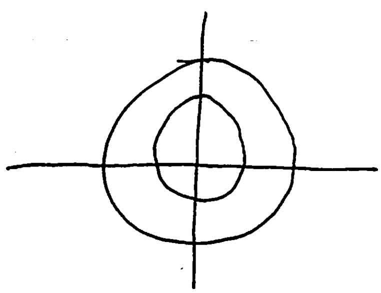
- If components of random vector are iid, call random vector iid.

EX: $\underline{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$ where w_i iid $\sim \mathcal{N}(0,1)$

$$f_{\underline{w}}(\underline{w}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-w_i^2/2} = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\underline{w}^T \underline{w}}{2}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\|w\|^2}{2}\right)$$

note $f_{\underline{w}}(w)$ depends only on $\underline{w}^T \underline{w} = \|\underline{w}\|^2$

So $f_{\underline{w}}(w)$ is spherically symmetric around origin



equal prob contours lie on concentric spheres around origin.

$$g_{\underline{w}}(\underline{s}) = E[e^{\underline{s}^T \underline{w}}] = E[e^{s_1 w_1 + \ldots + s_n w_n}] = E[\prod_{i=1}^{n} e^{s_i w_i}]$$

$$= \prod_{i=1}^{n} E[e^{s_i w_i}] = \prod_{i=1}^{n} e^{s_i^2/2} = \exp\left[\frac{\underline{s}^T \underline{s}}{2}\right]$$

$$= \exp\left[\frac{\|s\|^2}{2}\right]$$

Definition $\{z_1, \dots, z_n\}$ a set of jointly Gaussian r.v.'s and $\underline{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$ is a Gaussian random vector if for all real vectors $\underline{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}$, the linear combination

$$\underline{s}^T \underline{z} = \sum_{i=1}^{n} s_i z_i \text{ is a Gaussian r.v.}$$

Intuition: By CLT, sum of large # of small indep. r.v.'s is Gaussian. e.g. broadband noise passed through a narrowband linear filter: output at given time is Gaussian. Linear comb. of outputs at diff. times also sum of small set of underlying r.v.'s. Thus, sum is again ~ Gaussian. Thus, set of outputs at diff. times is jointly Gaussian.

- If $\{z_1, \dots, z_n\}$ JG, then z_i is Gaussian r.v. for each i. Take $s_i = 1$, $s_j = 0$ $j \neq i$. Any subset of $\{z_1, \dots, z_n\}$ JG.
- However, z_i's individually Gaussian $\not\Rightarrow$ $\{z_1, \dots, z_n\}$ JG.

  e.g. Let $z_1 \sim \mathcal{N}(0,1)$, $\underset{\substack{\uparrow \\ \text{indep of } z_1}}{X} = \begin{cases} +1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2. \end{cases}$ $\quad z_2 = z_1 X$

  Since Gaussian is symmetric about 0, z_2 is Gaussian. $\sim \mathcal{N}(0,1)$.

  But $z_1 + z_2 = 0$ w.p. $1/2$. Thus, $\{z_1, z_2\}$ not JG.

  (Also, $E[z_1 z_2] = E[z_1 z_1 X] = E[z_1^2] E[X] = 0 \Rightarrow z_1, z_2$ uncorrelated but not indep. However, if $\{z_1, z_2\}$ were JG, then uncorrelated $\Rightarrow$ indep.)

MGF of 0-mean Gaussian random vector $\underline{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$

$$g_{\underline{z}}(\underline{s}) = E[e^{\underline{s}^T \underline{z}}] = E[e^{\sum_i s_i z_i}]$$

Let $X = \underline{s}^T \underline{z}$. $\underline{z}$ Gaussian r. vector $\Rightarrow X$ Gaussian

$\underline{z}$ 0-mean $\Rightarrow X$ 0-mean.

Now $g_X(s) = \exp[s^2 \sigma_X^2/2] \Rightarrow g_X(1) = \exp(\sigma_X^2/2)$

$$g_{\underline{z}}(\underline{s}) = E[e^X] = \exp(\sigma_X^2/2)$$

$$\sigma_X^2 = E[X^2] = E[\underline{s}^T \underline{z}\, \underline{z}^T \underline{s}] = \underline{s}^T K_{\underline{z}}\, \underline{s}$$

$$\Rightarrow g_{\underline{z}}(\underline{s}) = \exp\left[\frac{\underline{s}^T K_{\underline{z}}\, \underline{s}}{2}\right]$$

- For non-zero mean JGRV. $\underline{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$ $m_i = E[u_i]$ $\underline{m} = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix}$

Let $\underline{u} = \underline{m} + \underline{z}$ where $\underline{z}$ is 0-mean JGRV.

$$g_{\underline{u}}(\underline{s}) = E[\exp(\underline{s}^T \underline{m} + \underline{s}^T \underline{z})] = \exp(\underline{s}^T \underline{m})\, g_{\underline{z}}(\underline{s})$$

Now $K_{\underline{u}} = K_{\underline{z}}$

$$= \exp\left(\underline{s}^T \underline{m} + \frac{\underline{s}^T K_{\underline{u}}\, \underline{s}}{2}\right) \quad (1)$$

completely specified by $\underline{m}$, $K_{\underline{u}}$

- Denote $\underline{u} \sim \mathcal{N}(\underline{m}, K_{\underline{u}})$
- conversely, can show if $\underline{u}$ has MGF in (1), then each linear comb. of components of $\underline{u}$ is Gaussian.

Thm: $\underline{u} \sim \mathcal{N}(\underline{m}, K_{\underline{u}})$ iff $g_{\underline{u}}(\underline{s})$ is given by (1).

Note if $w_1, \ldots, w_n$ iid $\sim \mathcal{N}(0,1)$, then $\underline{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$ is JGRV. $K_{\underline{w}} = I$, $m_i = 0$

$g_{\underline{w}}(\underline{s}) = \exp\left(\frac{\underline{s}^T \underline{s}}{2}\right)$ as before.

...nt Probability Densities for Gaussian Random Vectors

Let A be $n \times n$, $\underline{W}$ iid normalized Gaussian random vector

$\underline{Z} = A\underline{W}$ $\quad K_{\underline{Z}} = E[\underline{Z}\,\underline{Z}^T] = E[A\underline{W}\,\underline{W}^T A^T] = AA^T$ since $K_{\underline{W}} = I$

$$g_{\underline{Z}}(\underline{s}) = E[e^{\underline{s}^T \underline{Z}}] = E[\exp(\underline{s}^T A \underline{W})] = E[\exp((A^T\underline{s})^T \underline{W})]$$

$$= g_{\underline{W}}(A^T\underline{s}) = \exp\left[\frac{\underline{s}^T A A^T \underline{s}}{2}\right] = \exp\left[\frac{\underline{s}^T K_{\underline{Z}}\, \underline{s}}{2}\right]$$

Thus, $\underline{Z}$ is 0-mean GRV.

In fact, any 0-mean GRV can be represented in form $\underline{Z} = A\underline{W}$ for some A and iid normalized GRV $\underline{W}$.

Joint pdf of $\underline{Z} = A\underline{W}$:

Let $\underline{a}_1, \ldots, \underline{a}_n$ be n cols of A

Then
$$\underline{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} = \begin{bmatrix} \underline{a}_1 & \cdots & \underline{a}_n \end{bmatrix} \begin{bmatrix} W_1 \\ \vdots \\ W_n \end{bmatrix} = \sum_j \underline{a}_j W_j$$

- For any sample values $w_1, \ldots, w_n$ of $\underline{W}$, value of $\underline{Z}$ is $\underline{z} = \sum_j \underline{a}_j w_j$
- Assume A invertible, i.e. $\exists\, A^{-1}$ s.t. $AA^{-1} = A^{-1}A = I_n$

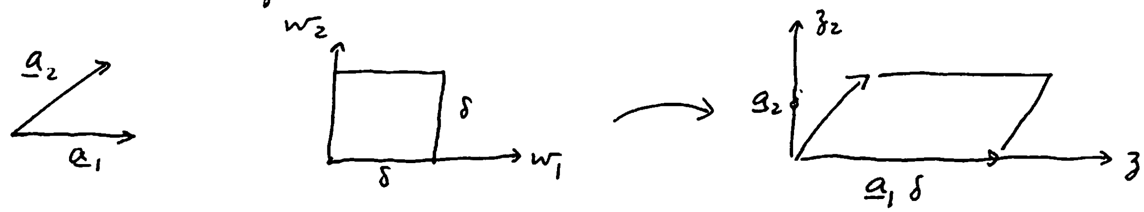
$\Rightarrow$ $\underline{a}_1, \ldots, \underline{a}_n$ forms basis for $\mathbb{R}^n$.

If A not invertible, then possible sample values for $\underline{Z}$ lie in proper subspace of $\mathbb{R}^n$ and $\underline{Z}$ would not have pdf.

- Consider small cube δ on a side of sample values for $\underline{W}$

i.e. Consider set B_δ of vectors for which $0 \le w_j \le \delta$ for $1 \le j \le n$

Set B'_δ of vectors $\underline{z} = A\underline{w}$ s.t. $\underline{w} \in B_\delta$



maps into parallelopiped whose sides are $\underline{a}_1\delta, \ldots, \underline{a}_n\delta$

$|\det(A)|$ = vol. of parall. with sides $\underline{a}_j$, $1 \le j \le n$

(See Strang) $\quad$ If $\det A = 0$, then singular case

Cube B_δ with vol. $\delta^n \xrightarrow{A}$ parall. with vol. $|\det A|\,\delta^n$

- Let z be sample value of $\underline{z}$

$\underline{w} = A^{-1} z$

Then $$\underbrace{f_{\underline{z}}(z)\,|dz|}_{\delta^n |\det A| \,=\, \text{vol. of parall.}} = \underbrace{f_{\underline{w}}(w)\,|d\underline{w}|}_{\delta^n = \text{vol. of } B_\delta} \qquad (2)$$

So $|dz| / |d\underline{w}| = |\det A|$

use in (2) and $f_{\underline{w}}(w) = f_{\underline{w}}(A^{-1} z)$

$$f_{\underline{z}}(z) = \frac{f_{\underline{w}}(A^{-1} z)}{|\det A|} = \frac{\exp\left[-\frac{1}{2} z^T (A^{-1})^T A^{-1} z\right]}{(2\pi)^{n/2}\,|\det A|}$$

Since $K_{\underline{z}} = AA^T$, $K_{\underline{z}}^{-1} = (A^{-1})^T A^{-1}$

$\det(K_{\underline{z}}) = \det(A) \det A^T = (\det A)^2 > 0$, ($A$ invertible)

Then $$f_{\underline{z}}(z) = \frac{\exp\left[-\frac{1}{2} z^T K_{\underline{z}}^{-1} z\right]}{(2\pi)^{n/2} \sqrt{\det K_{\underline{z}}}} \qquad (3)$$

- (3) has no meaning when $K_{\underline{z}}$ is singular, then $K_{\underline{z}}^{-1}$ does not exist. $A\underline{w}$ maps set of n-dim $\underline{w}$ to subspace of dim n. Some components of $\underline{z}$ can be expressed as lin comb of others. Define lin indep. components as random vector $\underline{z}'$ and work with this.

$P_{\underline{z}}(z)$ 0 outside of subspace, impulsive inside

- Now generalize: $\underline{u} = \underline{m} + A\underline{w} = \underline{m} + \tilde{\underline{u}}$ where $\tilde{\underline{u}} = A\underline{w}$

Assume $\det A \neq 0$

$$f_{\underline{u}}(u) = \exp \frac{\left[-\frac{1}{2}(u - m)^T K_{\underline{u}}^{-1} (u - m)\right]}{(2\pi)^{n/2} \sqrt{\det K_{\underline{u}}}}$$

$K_{\underline{u}} = E[\tilde{\underline{u}}\, \tilde{\underline{u}}^T] = K_{\tilde{\underline{u}}}$.

- general form of density for $\underline{u} \sim \mathcal{N}(m, K_{\underline{u}})$ if $\det(K_{\underline{u}}) \neq 0$.

end of Friday

- Assume $\underline{z}$ 0-mean for simplicity.
- $f_{\underline{z}}(\underline{z})$ depends on $K_{\underline{z}}$, not directly on A
- $\underline{z}^T K_{\underline{z}}^{-1} \underline{z}$ is quadratic in $\underline{z}$. $\{\underline{z} : \underline{z}^T K_{\underline{z}}^{-1} \underline{z} = c\}$ is an <u>ellipsoid</u> centered at origin

  $\Rightarrow$ contours of equal prob. forms set of concentric ellipsoids

- Axis of ellipsoids are eigenvectors of $K_{\underline{z}}$.
- <u>2-D example:</u> 0-mean GRV $\underline{z}$

$$\underline{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \qquad E[z_i^2] = \sigma_i^2 \qquad K_{\underline{z}} = \begin{bmatrix} \sigma_1^2 & k_{12} \\ k_{12} & \sigma_2^2 \end{bmatrix}$$

$E[z_1 z_2] = k_{12}$. Let ρ = normalized covariance $= k_{12}/(\sigma_1\sigma_2)$

For any joint distr. r.v.'s z_1, z_2, (HW).

$$(E[z_1 z_2])^2 \le E[z_1^2]\, E[z_2^2].$$

Thus, $|\rho| \le 1$.

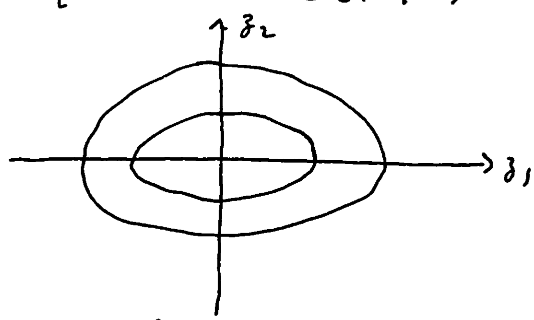
Now $\det(K_{\underline{z}}) = \sigma_1^2\sigma_2^2 - k_{12}^2 = \sigma_1^2\sigma_2^2(1-\rho^2) > 0$ iff $|\rho| < 1$

Then $$K_{\underline{z}}^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1/\sigma_1^2 & -\rho/(\sigma_1\sigma_2) \\ -\rho/(\sigma_1\sigma_2) & 1/\sigma_2^2 \end{bmatrix}$$

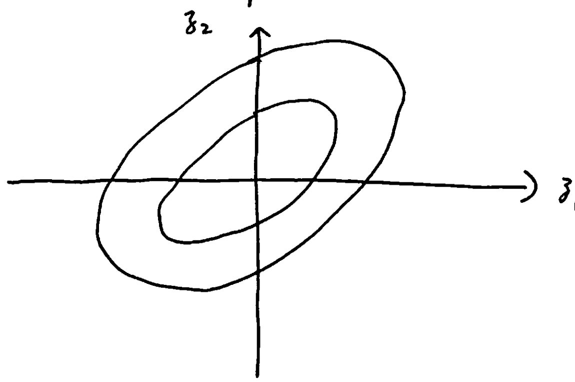
$$f_{\underline{z}}(\underline{z}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[\frac{-\left(\frac{z_1}{\sigma_1}\right)^2 + 2\rho\left(\frac{z_1}{\sigma_1}\right)\left(\frac{z_2}{\sigma_2}\right) - \left(\frac{z_2}{\sigma_2}\right)^2}{2(1-\rho^2)}\right]$$

If $\rho = 0$ and $\sigma_1 > \sigma_2$

contours:



If $\rho > 0$



- Much better to deal with vector notation!

Covariance Matrices

- following applies to Gaussian as well non-Gaussian case.
- K is covariance matrix if $\exists$ 0-mean RV $\underline{z}$ s.t. $K = E[\underline{z}\,\underline{z}^T]$
- $n \times n$ matrix K positive semidefinite (psd) if symmetric and $\underline{b}^T K \underline{b} \geq 0 \quad \forall\, \underline{b} \in \mathbb{R}^n$. It is positive definite (pd) if psd and $\underline{b}^T K \underline{b} > 0 \quad \forall\, \underline{b} \neq \underline{0}$

Properties of covariance matrices

1) Every cov. matrix K is psd.

symmetric: $K = E[\underline{z}\,\underline{z}^T]$ for 0-mean RV $\underline{z}$ $\quad K_{ij} = E[z_i z_j] = E[z_j z_i] = K_{ji} \quad \forall i,j$

Let $\underline{b} \in \mathbb{R}^n$, $X = \underline{b}^T \underline{z} \qquad 0 \leq E[X^2] = E[\underline{b}^T \underline{z}\,\underline{z}^T \underline{b}] = \underline{b}^T K \underline{b}$

2) Recall $\lambda \in \mathbb{C}$ is eigenvalue of K and $\underline{q} \neq \underline{0}$ is eigenvector of K if $K\underline{q} = \lambda \underline{q}$

- all eigenvalues of K are non-negative (positive) if K psd (pd)
- eigenvectors can be taken to be real
- eigenvectors of diff. eigenvalues are orthogonal
- If eigenvalue of multiplicity j, then it has j orthogonal eigenvectors
  $\Rightarrow$ can choose n orthogonal eigenvectors. Can normalize to orthonormal

3) If K psd, then $\exists$ orthonormal ~~matrix~~ (orthogonal) matrix whose cols $\underline{q}_1, \ldots, \underline{q}_n$ are orthonormal eigenvectors above s.t. $KQ = Q\Lambda \qquad \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \qquad \lambda_i = \text{eigenvalue for } \underline{q}_i$

$Q^T Q = I, \quad Q^{-1} = Q^T$.

Thus, $K = Q \Lambda Q^T$ ~~If K pd~~ If K pd, then $K^{-1} = Q \Lambda^{-1} Q^T$ since all $\lambda_i > 0$

4) $\det K = \prod_{i=1}^{n} \lambda_i$, $\quad \lambda_1, \ldots, \lambda_n$ eigenvalues.

If K pd (psd), $\det K > 0$ (≥ 0) since $\lambda_i > 0$ ($\lambda_i \geq 0$)

5) If K pd (psd), then $\exists$ unique pd (psd) square root matrix R s.t. $R^2 = K$.

$$R = Q \Lambda^{1/2} Q^T \qquad \Lambda^{1/2} = \begin{bmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{bmatrix}$$

6) K psd $\Rightarrow$ K a covariance matrix

K is cov. matrix of $\underline{z} = R\underline{w}$ where $R = Q\Lambda^{1/2}Q^T$

and $\underline{w}$ iid normalized GRV. $\quad K = E[R\underline{w}\,\underline{w}^T R] = R^2$

7) For any $n \times n$ matrix A, $\quad K = AA^T$ is a cov. matrix.

K is cov. matrix of $\underline{z} = A\underline{w} \qquad K = E[A\underline{w}\,\underline{w}^T A] = AA^T$

e.g. can take $A = R = Q\Lambda^{1/2}Q^T$

Corollary for Gaussian case:

For any cov. matrix K, a 0-mean GRV $\underline{z} \sim \mathcal{N}(0, K)$ exits and $\underline{z} = A\underline{w}$, where $K = AA^T$ and $\underline{w} \sim \mathcal{N}(\underline{0}, I_n)$

Geometry and Principle Axes

Let $\underline{z} \sim \mathcal{N}(0, K)$. where K is nonsingular.

$$f_{\underline{z}}(\underline{z}) = \frac{\exp[-\frac{1}{2}\underline{z}^T K_{\underline{z}}^{-1}\underline{z}]}{(2\pi)^{n/2}\sqrt{\det K_{\underline{z}}}}$$

Contour of equal prob density $= \{\underline{z} \mid \underline{z}^T K_{\underline{z}}^{-1}\underline{z} = c\}$ is an ellipsoid centered at origin.

Let $\quad K = Q\Lambda Q^T$ where Q is orthonormal, $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \ddots \lambda_n \end{bmatrix}$

$\underline{v} = Q^T \underline{z}$ also 0-mean GRV.

$$K_{\underline{v}} = E[\underline{v}\,\underline{v}^T] = E[Q^T\underline{z}\,\underline{z}^T Q] = Q^T K Q = \Lambda$$

$$f_{\underline{v}}(\underline{v}) = \frac{\exp[-\frac{1}{2}\underline{v}^T K_{\underline{v}}^{-1}\underline{v}]}{(2\pi)^{n/2}\sqrt{\det K_{\underline{v}}}} = \frac{\exp[-\sum_i v_i^2/(2\lambda_i)]}{(2\pi)^{n/2}\prod_i \sqrt{\lambda_i}}$$

$$= \prod_{i=1}^{n} \frac{\exp[-\frac{v_i^2}{2\lambda_i}]}{\sqrt{2\pi\lambda_i}}$$

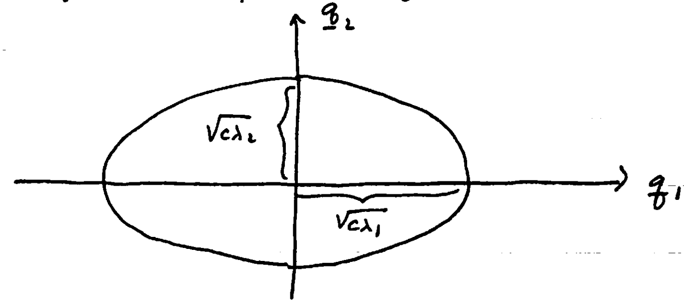
- $\underline{v} = Q^T \underline{z}$ is a change of basis (rotation since Q is orthonormal)
- In $\underline{v}$ - system

$$\{ \underline{v} : f_{\underline{v}}(\underline{v}) = c \} = \{ \underline{v} : \sum_i \frac{v_i^2}{\lambda_i} = c \} \text{ is an ellipsoid}$$

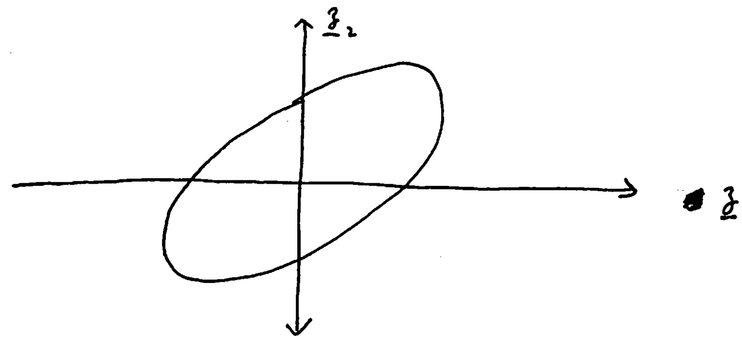
axes of ellipsoid are the eigenvectors $\underline{q}_1, \ldots, \underline{q}_n$ (cols of Q)

distance from origin to ellipsoid along q_i direction $= \sqrt{c\lambda_i}$

2-D example:



Back to $\underline{z}$ coordinates: $\underline{z} = Q\underline{v}$



Conditional Probabilities

X, Y jointly Gaussian 0-mean random variables with nonsingular cov. matrix.

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \quad \begin{array}{l} \leftarrow \text{use expression from before} \\ \leftarrow Y \sim N(0, \sigma_y^2) \end{array}$$

$$= \frac{1}{\sigma_x \sqrt{2\pi(1-\rho^2)}} \exp\left[\frac{-(x - \rho(\sigma_x/\sigma_y)y)^2}{2\sigma_x^2(1-\rho^2)} \right]$$

where $\rho = \frac{E[XY]}{\sigma_x \sigma_y}$

Given $Y = y$,

$f_{X|Y}(x|y) = N(\rho(\frac{\sigma_x}{\sigma_y})y, \; \sigma_x^2(1-\rho^2))$!

- Given $Y = y$, X is conditionally Gaussian, with mean $\rho\left(\frac{\sigma_X}{\sigma_Y}\right) y$ linear in y
  the fluctuation $\tilde{X}$ of X (with mean $\rho\left(\frac{\sigma_X}{\sigma_Y}\right) y$ removed)
  has same density for all y. i.e. $\tilde{X} \sim \mathcal{N}(0, \sigma_X^2(1-\rho^2))$
- very important for estimation
- Generalize to higher dimensions:

$$\underline{u} = \begin{bmatrix} u_1 \\ \vdots \\ \vdots \\ u_{n+m} \end{bmatrix} \quad \text{GRV} \quad = \begin{bmatrix} X_1 \\ \vdots \\ X_m \\ Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \underline{X} \\ \underline{Y} \end{bmatrix} \qquad \underline{X}, \underline{Y} \text{ JGRV's.}$$

If K_u is nonsingular, $\underline{X}, \underline{Y}$ are jointly non-singular.

Assume $\underline{X}, \underline{Y}$ JGRV, Jointly non-sing, 0-mean

Note: $K_u = \begin{bmatrix} K_X & K_{XY} \\ K_{XY}^T & K_Y \end{bmatrix}$ $\qquad K_X = E[\underline{X}\,\underline{X}^T]$, $\quad K_Y = E[\underline{Y}\,\underline{Y}^T]$, $\quad K_{XY} = E[\underline{X}\,\underline{Y}^T]$

$K_u^{-1} = \begin{bmatrix} B & C \\ C^T & D \end{bmatrix}$ $\qquad K_u$ symmetric $\Rightarrow$ K_u^{-1} symmetric

B, D also symmetric

can also show K_X, K_Y, B, D nonsing.

Now

$$f_{\underline{X},\underline{Y}}(\underline{x}, \underline{y}) = \frac{\exp\left[-\frac{1}{2}[\underline{x}^T\ \underline{y}^T]\, K_u^{-1} \begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix}\right]}{(2\pi)^{(n+m)/2}\sqrt{\det K_u}}$$

$$= \frac{\exp\left[-\frac{1}{2}\left(\underline{x}^T B \underline{x} + \underline{y}^T C^T \underline{x} + \underline{x}^T C \underline{y} + \underline{y}^T D \underline{y}\right)\right]}{(2\pi)^{(n+m)/2}\sqrt{\det K_u}}$$

Note $\underline{x}$ appears only in first 3 terms in exponent

Now $\quad f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \dfrac{f_{\underline{X},\underline{Y}}(\underline{x}, \underline{y})}{f_{\underline{Y}}(\underline{y})}$ $\leftarrow$ does not involve $\underline{x}$

$$= \underbrace{g(\underline{y})}_{\text{some fn. of } \underline{y}} \exp\left[-\frac{(\underline{x}^T B \underline{x} + \underline{y}^T C^T \underline{x} + \underline{x}^T C \underline{y})}{2}\right]$$

Complete square in exponent around B:

$$f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = g(\underline{y}) \exp\left[\frac{-(\underline{x} + B^{-1}C\underline{y})^T B(\underline{x} + B^{-1}C\underline{y}) + \underline{y}^T C^T B^{-1} C \underline{y}}{2}\right]$$

Last term does not depend on $\underline{x}$, absorb into $g(\underline{y})$:

$$f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = g'(\underline{y}) \exp\left[-\tfrac{1}{2}(\underline{x} + B^{-1}C\underline{y})^T B(\underline{x} + B^{-1}C\underline{y})\right]$$

Since $f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y})$ is a pdf for each $\underline{y}$, $g'(\underline{y})$ must be s.t.

$$f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \frac{\exp\left[-\tfrac{1}{2}(\underline{x} + B^{-1}C\underline{y})^T B(\underline{x} + B^{-1}C\underline{y})\right]}{(2\pi)^{m/2}\sqrt{\det B^{-1}}}$$

- Thus, for any given $\underline{y}$, $f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = n(-B^{-1}C\underline{y}, B^{-1})$

  Covariance B^{-1} does not depend on $\underline{y}$

  Mean $-B^{-1}C\underline{y}$ linear in $\underline{y}$

- Given $\underline{Y} = \underline{y}$, $\underline{X} = -B^{-1}C\underline{y} + \underline{V}$

  where $\underline{V}$ is a Gaussian fluctuation of covariance B^{-1} indep of $\underline{y}$

  $\underline{V}$ indep of $\underline{Y}$

- Thus, $\underline{X} = G\underline{Y} + \underline{V}$ $\quad \underline{Y}, \underline{V}$ indep.

  $G = -B^{-1}C$, $\quad \underline{V} \sim n(0, B^{-1})$.

  $\uparrow$ <u>innovation</u> (part of $\underline{X}$ indep of $\underline{Y}$)

  or <u>noise term</u>

Call $K_{\underline{V}} = B^{-1}$ <u>conditional</u> <u>covariance</u> of $\underline{X}$ given any sample value $\underline{y}$ of $\underline{Y}$

Unconditional covariance of $\underline{X} = K_X$ upper left block of K_u

conditional " of $\underline{X} = K_V = B^{-1}$ " of K_u^{-1}

– Now $\quad \underline{X} = G\underline{Y} + \underline{V} \qquad \underline{V}, \underline{Y}$ indep.

$$K_{XY} = E[\underline{X}\,\underline{Y}^T] = E[G\underline{Y}\,\underline{Y}^T + \underline{V}\,\underline{Y}^T] = G K_{\underline{Y}}$$

$$K_X = E[(G\underline{Y}+\underline{V})(G\underline{Y}+\underline{V})^T] = G K_Y G^T + K_V$$

$$\Rightarrow \qquad G = K_{XY} K_Y^{-1}$$

$$K_V = K_X - G K_Y G^T = K_X - K_{XY} K_Y^{-1} K_{XY}^T$$

$$f_{\underline{X}|\underline{Y}}(\underline{x}|\underline{y}) = \frac{\exp[-\frac{1}{2}(\underline{x} - G\underline{y})^T K_V^{-1} (\underline{x} - G\underline{y})]}{(2\pi)^{m/2}\sqrt{\det K_V}}$$

– $\quad B = K_V^{-1} = (K_X - K_{XY} K_Y^{-1} K_{XY}^T)^{-1}$

$$C = -BG = -K_V^{-1} G$$

$$= -K_V^{-1} K_{XY} K_Y^{-1}$$

– Using symmetry bet X and Y, can also write

$$\underline{Y} = H\underline{X} + \underline{Z}, \qquad \underline{X}, \underline{Z} \text{ indep}$$

$$H = -D^{-1} C^T, \qquad \underline{Z} \sim \mathcal{N}(0, D^{-1}).$$

$$K_{XY} = K_X H^T, \qquad K_Y = H K_X H^T + K_Z$$

$$K_Z = K_Y - K_{XY}^T K_X^{-1} K_{XY}.$$

$$p_{\underline{Y}|\underline{X}}(\underline{y}|\underline{x}) = \frac{\exp[-\frac{1}{2}(\underline{y} - H\underline{x})^T K_Z^{-1} (\underline{y} - H\underline{x})]}{(2\pi)^{n/2}\sqrt{\det K_Z}}$$

$$D = (K_Y - K_{XY}^T K_X^{-1} K_{XY})^{-1}$$

$$C^T = -K_Z^{-1} K_{XY}^T K_X^{-1}$$

– Start with $\underset{\substack{\uparrow \\ \text{observation}}}{\underline{Y}} = H\underline{X} + \underset{\substack{\uparrow \\ \text{noise}}}{\underline{Z}}$

find G and K_V s.t. $\underline{X} = G\underline{Y} + \underline{V}$ $\Big\}$ estimation

Can show $\quad G = K_V H^T K_Z^{-1}$

$$K_V^{-1} = K_X^{-1} + H^T K_Z^{-1} H$$