ii. Directional preference

Suppose we are given the covariance matrix $\underline{K}_X$ of some vector $\underline{X}(u)$ and would like to project this vector on some unit-length vector $\underline{b}$ (i.e. $sum_{i=1}^{n} b_i^2 = 1$). The projection is in the inner product sense

$$Y(u) = (\underline{b}, \underline{X}(u)) = \underline{X}(u)^T \underline{b}$$

Lets assume that $\underline{m}_X = \underline{0}$ and let's calculate the variance of $Y(u)$:

$$\begin{aligned} var\{Y(u)\} &= {\sigma_Y}^2 = \mathcal{E}\{Y^2(u)\} = \\ &= \mathcal{E}\{Y(u)Y(u)\} = \mathcal{E}\{\underline{b}^T \underline{X}(u)\underline{X}(u)^T \underline{b}\} \\ &= \underline{b}^T \underline{K}_X \underline{b} \end{aligned}$$

i.e. the variance of $Y(u)$ is a quadratic functional of the $\underline{b}_i$'s. "Directional preference" translates to finding those directions $\underline{b}$ where the variance is highest (or lowest). This is an optimization problem where we want to maximize the above quadratic form subject to the norm constraint. To do that, project $\underline{b}$ on the $\underline{e}_i$'s which form an orthonormal basis.

$$\underline{b} = \sum_i b_i \underline{e}_i$$

so that $\sum_{i=1}^{n} b_i^2 = 1$.
We can also rework the quadratic form as follows:

$$\begin{aligned} {\sigma_Y}^2 &= \underline{b}^T \underline{K}_X \underline{b} = \sum_i b_i {\underline{e}_i}^T \underline{K}_X (\sum_j b_j \underline{e}_j) \\ &= \sum_i \sum_j b_i b_j {\underline{e}_i}^T \underline{K}_X \underline{e}_j \end{aligned}$$

But $\underline{K}_X \underline{e}_j = \lambda_j \underline{e}_j$ because all the $\underline{e}_j$'s are eigenvectors of $\underline{K}_X$. So

$$\begin{aligned} {\sigma_Y}^2 &= \sum_i \sum_j b_i b_j {\underline{e}_i}^T \lambda j \underline{e}_j \\ &= \sum_i \sum_j \lambda_j b_i b_j {\underline{e}_i}^T \underline{e}_j \end{aligned}$$

or

$${\sigma_Y}^2 = \sum_i \lambda_i b_{i2}$$

An equivalent problem is the following: Let $u_i \stackrel{\text{def}}{=} {b_i}^2$ and we want to

maximize $\sum_{i=1}^{n} \lambda_i u_i = U$ subject to the linear constraint $\sum_{i=1}^{n} u_i = 1$ and $u_i \geq 0, \lambda_i \geq 0$ This is a standard linear programming problem.

Example :

For $\lambda_2 > \lambda_1$ the optimal solution is $u_1 = 0$, $u_2 = 1$. The general solution is to choose $u_j = 1$ where $\lambda_j = max\{\lambda_i\}$ and $u_i = 0$ for $i \neq j$.
i.e. since $b_i{}^2 = 1$ it follows that $b_j = \pm 1$, $b_i = 0$, $i \neq j$.
The resulting variance is the maximum eigenvalue

$$\sigma_Y{}^2 = \lambda_j$$

But recall that

$$\underline{b} = \sum_{i=1}^{n} b_i \underline{e}_i \Rightarrow \underline{b}_{max} = (\underline{e}_{imax})$$

i.e. the direction that maximizes the variance is the direction of the *eigenvector* corresponding to the largest eigenvalue. Another kind of use of the correlation (or covariance) matrix is in deriving certain important *inequalities*, which are useful in probability in bounding probabilities of events. Inequalities are convenient to work with whenever *exact* calculation of probabilities is prohibitive in terms of complexity.

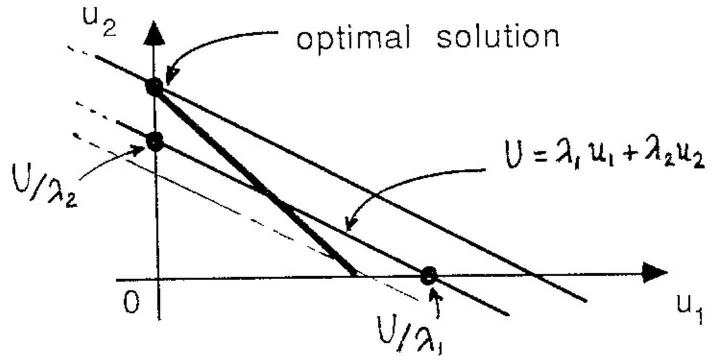
The Chebyshev Inequality

One such widely used inequality is the *Chebychev* inequality, extended to random vectors. The goal is to upperbound the probability that the length of a vector exceeds a certain number

$$Pr\{|\underline{X}(u)| > \epsilon\} \leq g(\epsilon)$$

Suppose we are given $\underline{R}_X$. Then

$$\begin{aligned} tr(\underline{R}_X) &= R_X(1,1) + \cdots + R_X(n,n) \\ &= \sum_{i=1}^{n} \mathcal{E}\{X_i{}^2(u)\} \\ &= \mathcal{E}\{|\underline{X}(u)|^2\} = \underbrace{\int \cdots \int}_{n-fold} |\underline{x}|^2 f_{\underline{X}}(\underline{x}) d\underline{x} \end{aligned}$$



u_2
optimal solution
$U = \lambda_1 u_1 + \lambda_2 u_2$
U/λ_2
0
u_1
U/λ_1

F 2-2

$$
\begin{aligned}
&= \underbrace{\int \cdots \int}_{|\underline{X}| \le \epsilon} |\underline{x}|^2 f_{\underline{X}}(\underline{x}) d\underline{x} + \underbrace{\int \cdots \int}_{|\underline{X}| > \epsilon} |\underline{x}|^2 f_{\underline{X}}(\underline{x}) d\underline{x} \\
&\ge \underbrace{\int \cdots \int}_{|\underline{X}| > \epsilon} |\underline{x}|^2 f_{\underline{X}}(\underline{x}) d\underline{x}
\end{aligned}
$$

because the first term is ommited, being nonnegative. We now observe that

$$
\begin{aligned}
tr(\underline{R}_X) &\ge \epsilon^2 \underbrace{\int \cdots \int}_{|\underline{X}| > \epsilon} f_{\underline{X}}(\underline{x}) d\underline{x} \\
&\ge \epsilon^2 Pr\{|\underline{X}| > \epsilon\}
\end{aligned}
$$

$$
\Rightarrow \boxed{Pr\{|\underline{X}| > \epsilon\} \le \frac{tr(\underline{R}_X)}{\epsilon^2}}
$$

This is the vector version of the Chebychev inequality. Notice that as ϵ increases, it is less likely to get vectors with large length. The inequality is general and does not require that $\underline{X}$ have zero mean.

We next examine in more detail the scalar case, where $\underline{X}(u) = x(u)$

$$
Pr\{| X(u) | > \epsilon\} \le \frac{\mathcal{E}\{X^2\}}{\epsilon^2}
$$

Suppose that $m_X \ne 0$ and define the normalized r.v.

$$
X' \stackrel{\text{def}}{=} \frac{X - m_X}{\sigma_X}
$$

Then

$$
\mathcal{E}\{X'\} = 0 \ , \ var\{X'\} = 1
$$

Applying the Chebychev inequality for x' we get

$$
\begin{aligned}
& Pr\{|X'| > \epsilon\} \le \frac{1}{\epsilon^2} \\
\Rightarrow \ & Pr\left[\left|\frac{X - m_X}{\sigma_X}\right| > \epsilon\right] \le \frac{1}{\epsilon^2} \\
\Rightarrow \ & Pr\{|X - m_X| > \epsilon\sigma_X\} \le \frac{1}{\epsilon^2} \\
\Rightarrow \ & Pr\{X > m_X + \epsilon\sigma_X \cup X < m_X - \epsilon\sigma_X\} \le \frac{1}{\epsilon^2}
\end{aligned}
$$

Regarding the tightness of the inequality we look at a specific example and let us choose the Gaussian case i.e.

$$X \sim G(m_X \ ; \ \sigma_X{}^2)$$

and also define the Gaussian tail integral

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

so the cumulative distribution function is

$$1 - Q(x) = F_X(x)$$

Then

$$Pr\{|X - m_X| > \epsilon\sigma_X\} = Pr\{|X'| > \epsilon\} = 2Q(\epsilon)$$

Another important application of the Chebychev inequality is the "weak

F2-3

law of large numbers ". Suppose that we have a sequence of independent, identically distributed (i.i.d.) random variables X_i , $i = 1, 2, \ldots$, all with mean m_X and variance $\sigma_X{}^2$. Consider the first N. Based on those we create the " sample mean ", a r.v.

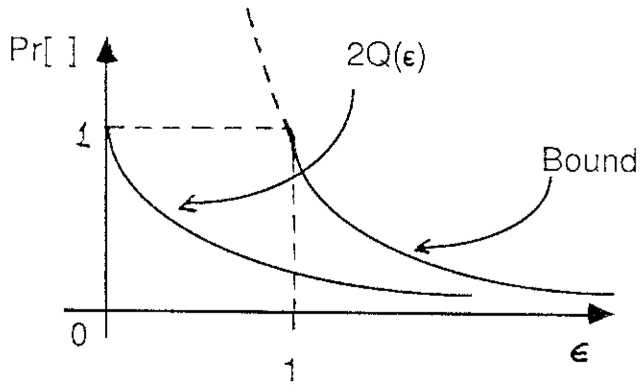
$$m(u) = \frac{1}{N}\sum_{i=1}^{N} X_i$$

The *ensemble* average (mean) of this " time average " is

$$\mathcal{E}\{m(u)\} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{E}\{X_i\} = m_X$$

Regards the variance of $m(u)$,

$$\begin{aligned}
\sigma_m{}^2 &= \mathcal{E}\left\{(m(u) - \mathcal{E}\{m(u)\})^2\right\} \\
&= \mathcal{E}\left\{\left[\frac{1}{N}\sum_{i=1}^{N} X_i - \frac{N m_X}{N}\right]^2\right\}
\end{aligned}$$



F2-3

$$= \mathcal{E}\left\{\left[\frac{1}{N}\sum_{i=1}^{N}(X_i - m_X)\right]^2\right\}$$

$$\Rightarrow \quad {\sigma_m}^2 = \frac{1}{N^2}(N{\sigma_X}^2)$$

$$\Rightarrow \quad \boxed{{\sigma_m}^2 = \frac{{\sigma_X}^2}{N}}$$

From Chebyshev inequality we conclude that

$$\Rightarrow Pr\{\mid m(u) - m_X \mid \geq \epsilon\} \leq \frac{{\sigma_X}^2}{N\epsilon^2}$$

We can see now that as

$$N \to \infty \ , \ \frac{{\sigma_X}^2}{N\epsilon^2} \to 0$$

$$\Rightarrow \lim_{N\to\infty} Pr\left\{\left|\frac{1}{N}\sum_{i=1}^{N} X_i - m_X\right| \geq \epsilon\right\} = 0$$

for every ϵ, which is the *weak law of large numbers.*

2.3 The whitening concept

This is the converse to the factorization or spectral shaping or coloring problem. There, given $\underline{W}(u)$ we wanted to find $\underline{H}$ such that $\underline{H}\underline{W}(u)$ had a given $\underline{K}_X$. Here we are given a random vector $\underline{X}(u)$ with some mean $\underline{m}_X$ and covariance $\underline{K}_X$ and we would like to find a linear transformatin $\underline{G}$ such that the output is a white vector $\underline{W}(u)$.

i. $\underline{m}_X = \underline{0}$

F 2-4

ii. $\underline{m}_X \neq \underline{0}$. In this case we must subtract the mean.

F 2-5