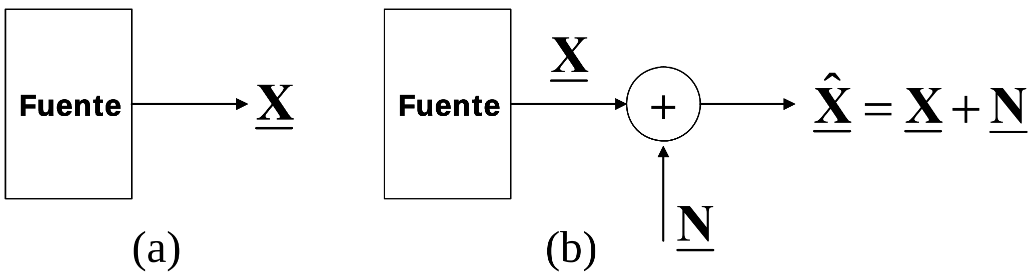
REDUCCIÓN DE RANGO

Una fuente genera vectores de datos reales aleatorios $\underline{\mathbf{X}} = [x_1, x_2, ..., x_N]^T$, estacionarios en sentido amplio, de acuerdo a la figura (a). La descomposición de la matriz de autocorrelación, en autovectores y autovalores, viene dada por $\underline{\underline{\mathbf{R}}} = E\left[\underline{\mathbf{X}}\underline{\mathbf{X}}^T\right] = \underline{\underline{\mathbf{Q}}}\,\underline{\underline{\mathbf{\Lambda}}}\,\underline{\underline{\mathbf{Q}}}^T$, donde la matriz de autovectores $\underline{\underline{\mathbf{Q}}} = \left[\underline{\mathbf{q}}_1, \underline{\mathbf{q}}_2, ..., \underline{\mathbf{q}}_N\right]$ es unitaria y la matriz diagonal de autovalores viene dada por $\underline{\underline{\mathbf{\Lambda}}} = diag[\lambda_1, \lambda_2, ..., \lambda_N]$. Los autovalores están ordenados de mayor a menor, es decir, $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$.

Las observaciones $\hat{\underline{\mathbf{X}}} = \underline{\mathbf{X}} + \underline{\mathbf{N}}$ son versiones de los datos $\underline{\mathbf{X}}$, degradadas por un término de ruido $\underline{\mathbf{N}}$ según se indica en la figura (b). El término de ruido $\underline{\mathbf{N}}$ es blanco, de media nula y con matriz de autocorrelación $E\left[\underline{\mathbf{N}}\underline{\mathbf{N}}^T\right] = \sigma_N^2 \underline{\underline{\mathbf{I}}}_{NxN}$.



(a) (b)

Una técnica frecuentemente utilizada en Procesado de la Señal es la denominada de '*Reducción de Rango*' ('*Rank Reduction*'). Dicha técnica consiste en lo siguiente. Considere el siguiente vector transformado:

$$\hat{\underline{\mathbf{X}}}_r = \underline{\underline{\mathbf{P}}}_r \left(\underline{\mathbf{X}} + \underline{\mathbf{N}}\right)$$

donde $\underline{\underline{\mathbf{P}}}_r = \underline{\underline{\mathbf{Q}}}_r \underline{\underline{\mathbf{Q}}}_r^T$ es una matriz de proyección, siendo la matriz $\underline{\underline{\mathbf{Q}}}_r = \left[\underline{\mathbf{q}}_1, \underline{\mathbf{q}}_{2,...,} \underline{\mathbf{q}}_r\right] \quad r \leq N$ una matriz compuesta por los primeros r autovectores (columna) de la matriz de autocorrelación $\underline{\underline{\mathbf{R}}}$.

Algunas propiedades de la matriz de proyección $\underline{\underline{\mathbf{P}}}_r$ son las siguientes:

- $\underline{\underline{\mathbf{P}}}_r$ y $\left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)$ son matrices simétricas.
- $\underline{\underline{\mathbf{P}}}_r$ y $\left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)$ son matrices idempotentes, es decir, $\underline{\underline{\mathbf{P}}}_r^2 = \underline{\underline{\mathbf{P}}}_r$ y $\left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)^2 = \left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)$.

Así como, también:

$$\underline{\underline{\mathbf{Q}}}^T \underline{\underline{\mathbf{P}}}_r \underline{\underline{\mathbf{Q}}} = \underline{\underline{\mathbf{I}}}_r \quad con \quad \underline{\underline{\mathbf{I}}}_r = diag[\underbrace{1,1,...,1}_{r},\underbrace{0,0,...,0}_{(N-r)}]$$

El ejercicio trata de estudiar la existencia de valores de r para los que es posible afirmar que la proyección de la observación $\underline{\hat{\mathbf{X}}}_r = \underline{\underline{\mathbf{P}}}_r\left(\underline{\mathbf{X}}+\underline{\mathbf{N}}\right)$ es una estimación más fiel del valor real $\underline{\mathbf{X}}$ que la propia observación $\underline{\hat{\mathbf{X}}} = \underline{\mathbf{X}}+\underline{\mathbf{N}}$.

Se pide conteste a las siguientes preguntas:

1.- Obtenga la expresión de la matriz de autocorrelación de las observaciones $\underline{\hat{\mathbf{X}}} = \underline{\mathbf{X}}+\underline{\mathbf{N}}$.

Dados los términos estadísticamente independientes y de media cero, tenemos que:

$$\underline{\underline{\mathbf{R}}}_{\hat{x}\hat{x}} = E\left[\underline{\hat{\mathbf{X}}}\underline{\hat{\mathbf{X}}}^T\right] = E\left[\underline{\mathbf{X}}\underline{\mathbf{X}}^T\right] + E\left[\underline{N}\,\underline{N}^T\right] = \underline{\underline{\mathbf{R}}} + \sigma_N^2 \mathbf{I}_{NxN}$$

2. Obtenga la expresión de la matriz de autocorrelación del vector de observaciones proyectado $\underline{\hat{\mathbf{X}}}_r = \underline{\underline{\mathbf{P}}}_r\left(\underline{\mathbf{X}}+\underline{\mathbf{N}}\right)$.

Si procedemos como en (1.), de nuevo obtenemos:

$$\underline{\underline{\mathbf{R}}}_{\hat{x}_r\hat{x}_r} = E\left[\underline{\hat{\mathbf{X}}}_r\underline{\hat{\mathbf{X}}}_r^T\right] = \underline{\underline{\mathbf{P}}}_r E\left[\underline{\mathbf{X}}\underline{\mathbf{X}}^T\right]\underline{\underline{\mathbf{P}}}_r^T + \underline{\underline{\mathbf{P}}}_r E\left[\underline{N}\,\underline{N}^T\right]\underline{\underline{\mathbf{P}}}_r^T = \underline{\underline{\mathbf{P}}}_r\underline{\underline{\mathbf{R}}}\underline{\underline{\mathbf{P}}}_r^T + \sigma_N^2\underline{\underline{\mathbf{P}}}_r\underline{\underline{\mathbf{P}}}_r^T$$

Podemos simplificar la expresión anterior si observamos el enunciado. Vemos que

$$\underline{\underline{\mathbf{Q}}}^T\underline{\underline{\mathbf{P}}}_r\underline{\underline{\mathbf{Q}}} = \underline{\underline{\mathbf{I}}}_r \Rightarrow \underline{\underline{\mathbf{Q}}}\underline{\underline{\mathbf{Q}}}^T\underline{\underline{\mathbf{P}}}_r\underline{\underline{\mathbf{Q}}} = \underline{\underline{\mathbf{Q}}}\underline{\underline{\mathbf{I}}}_r \Rightarrow \underline{\underline{\mathbf{P}}}_r\underline{\underline{\mathbf{Q}}} = \underline{\underline{\mathbf{Q}}}\underline{\underline{\mathbf{I}}}_r$$

De modo que

$$\underline{\underline{\mathbf{P}}}_r\underline{\underline{\mathbf{R}}}\underline{\underline{\mathbf{P}}}_r^T = \underline{\underline{\mathbf{P}}}_r\underline{\underline{\mathbf{Q}}}\underline{\underline{\mathbf{\Lambda}}}\underline{\underline{\mathbf{Q}}}^T\underline{\underline{\mathbf{P}}}_r^T = \left(\underline{\underline{\mathbf{P}}}_r\underline{\underline{\mathbf{Q}}}\right)\underline{\underline{\mathbf{\Lambda}}}\left(\underline{\underline{\mathbf{Q}}}^T\underline{\underline{\mathbf{P}}}_r^T\right) = \left(\underline{\underline{\mathbf{Q}}}\underline{\underline{\mathbf{I}}}_r\right)\underline{\underline{\mathbf{\Lambda}}}\left(\underline{\underline{\mathbf{Q}}}\underline{\underline{\mathbf{I}}}_r\right)^T = \underline{\underline{\mathbf{Q}}}\underline{\underline{\mathbf{I}}}_r\underline{\underline{\mathbf{\Lambda}}}\underline{\underline{\mathbf{I}}}_r\underline{\underline{\mathbf{Q}}}^T = \sum_{i=1}^{r}\lambda_i\underline{\mathbf{q}}_i\underline{\mathbf{q}}_i^T$$

El resultado es que los datos proyectados sólo contribuyen a la matriz de correlación con los 'r' modos principales.

3. Obtenga la expresión de la relación Señal-Ruido (SNR) de las observaciones *en función de los autovalores* de la matriz de autocorrelación $\underline{\underline{\mathbf{R}}}$ *y de* σ_N^2. Tenga en cuenta que:

$$SNR = \frac{tr\left(E\left[\underline{\mathbf{X}}\,\underline{\mathbf{X}}^T\right]\right)}{tr\left(E\left[\underline{\mathbf{N}}\,\underline{\mathbf{N}}^T\right]\right)}$$

donde $tr(...)$ indica la *traza* de una matriz.

La traza de una matriz de correlación es un invariante. De todos modos, podemos proceder formalmente:

$$SNR = \frac{tr\left(E\left[\underline{\mathbf{X}}\,\underline{\mathbf{X}}^T\right]\right)}{tr\left(E\left[\underline{\mathbf{N}}\,\underline{\mathbf{N}}^T\right]\right)} = \frac{tr\left(\underline{\underline{\mathbf{R}}}\right)}{tr\left(\sigma_n^2\underline{\underline{\mathbf{I}}}_{NxN}\right)}\left(= \frac{\cancel{N}\sigma_x^2}{\cancel{N}\sigma_n^2} = SNR\right) = \frac{tr\left(\underline{\underline{\mathbf{Q}}}\underline{\underline{\mathbf{\Lambda}}}\underline{\underline{\mathbf{Q}}}^T\right)}{N\sigma_n^2} = \frac{tr\left(\underline{\underline{\mathbf{\Lambda}}}\underline{\underline{\mathbf{Q}}}^T\underline{\underline{\mathbf{Q}}}\right)}{N\sigma_n^2} =$$

$$= \frac{tr\left(\underline{\underline{\mathbf{\Lambda}}}\right)}{N\sigma_n^2} = \frac{1}{N\sigma_n^2}\sum_{i=1}^{N}\lambda_i = \frac{\overline{\lambda}}{\sigma_n^2} \quad con: \quad \overline{\lambda} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i$$

4. Demuestre que el *sesgo* o *bias* promedio cuando tomamos la proyección $\underline{\hat{\mathbf{X}}}_r = \underline{\underline{\mathbf{P}}}_r\left(\underline{\mathbf{X}}+\underline{\mathbf{N}}\right)$ como una estimación de $\underline{\mathbf{X}}$ vendrá dado por:

$$b_r^2 = \frac{1}{N} tr\left(E\left[\left(\underline{\mathbf{X}} - E\left[\underline{\hat{\mathbf{X}}}_r\right]\right)\left(\underline{\mathbf{X}} - E\left[\underline{\hat{\mathbf{X}}}_r\right]\right)^T \right] \right) = \frac{1}{N} \sum_{i=r+1}^{N} \lambda_i$$

Cuando tomamos $\underline{\hat{\mathbf{X}}}_r$ como estimador de $\underline{\mathbf{X}}$ introducimos un sesgo. Veamos, para una observación dada a la salida de la fuente $\underline{\mathbf{X}}$, tenemos que:

$$E\left[\underline{\hat{\mathbf{X}}}_r\right] = E\left[\underline{\underline{\mathbf{P}}}_r\left(\underline{\mathbf{X}} + \underline{\mathbf{N}}\right)\right] = \underline{\underline{\mathbf{P}}}_r \underline{\mathbf{X}}$$

Para manejar una notación más compacta, definimos el vector diferencia:

$$\underline{\boldsymbol{\alpha}} = \underline{\mathbf{X}} - E\left[\underline{\hat{\mathbf{X}}}_r\right] = \left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)\underline{\mathbf{X}}$$

El término de bias (cuadrático) será la traza normalizada de la matriz de covarianza asociada:

$$E\left[\underline{\boldsymbol{\alpha}}\underline{\boldsymbol{\alpha}}^T\right] = \left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right) E\left[\underline{\mathbf{X}}\underline{\mathbf{X}}^T\right]\left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)^T = \left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)\underline{\underline{\mathbf{R}}}\left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)^T$$

Finalmente:

$$tr\left\{E\left[\underline{\boldsymbol{\alpha}}\underline{\boldsymbol{\alpha}}^T\right]\right\} = tr\left\{\left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)\underline{\underline{\mathbf{R}}}\left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)^T\right\} = tr\left\{\underline{\underline{\mathbf{R}}}\left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)^T\left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)\right\} =$$

$$= tr\left\{\underline{\underline{\mathbf{R}}}\left(\underline{\underline{\mathbf{I}}}_{NxN} - \underline{\underline{\mathbf{P}}}_r\right)\right\} = tr\left\{\underline{\underline{\mathbf{R}}}\right\} - tr\left\{\underline{\underline{\mathbf{R}}}\underline{\underline{\mathbf{P}}}_r\right\} = \sum_{i=1}^{N}\lambda_i - tr\left\{\underline{\underline{\mathbf{Q}}}\underline{\underline{\boldsymbol{\Lambda}}}\underline{\underline{\mathbf{Q}}}^T\underline{\underline{\mathbf{P}}}_r\right\} =$$

$$= \sum_{i=1}^{N}\lambda_i - tr\left\{\underline{\underline{\boldsymbol{\Lambda}}}\underline{\underline{\mathbf{Q}}}^T\underline{\underline{\mathbf{P}}}_r\underline{\underline{\mathbf{Q}}}\right\} = \sum_{i=1}^{N}\lambda_i - tr\left\{\underline{\underline{\boldsymbol{\Lambda}}}\underline{\underline{\mathbf{I}}}_r\right\} = \sum_{i=1}^{N}\lambda_i - \sum_{i=1}^{r}\lambda_i = \sum_{i=r+1}^{N}\lambda_i$$

y:

$$b_r^2 = \frac{1}{N} tr\left(E\left[\underline{\boldsymbol{\alpha}}\underline{\boldsymbol{\alpha}}^T\right]\right) = \frac{1}{N}\sum_{i=r+1}^{N}\lambda_i$$

La interpretación del resultado indica que como consecuencia de proyectar los datos, la observación queda sesgada por la ausencia de los modos suprimidos en la proyección. Dado que los autovalores son no negativos, el sesgo decrece con 'r'. Como cabe esperar, cuando 'r=N', el sesgo (cuadrático) será nulo dado que no se suprime ningún modo de la señal.

5. Demuestre que la *varianza* cuando tomamos la proyección $\underline{\hat{\mathbf{X}}}_r = \underline{\underline{\mathbf{P}}}_r\left(\underline{\mathbf{X}} + \underline{\mathbf{N}}\right)$ como una estimación de $\underline{\mathbf{X}}$ vendrá dado por:

$$\sigma_r^2 = \frac{1}{N} tr\left(E\left[\left(\underline{\hat{\mathbf{X}}}_r - E\left[\underline{\hat{\mathbf{X}}}_r\right]\right)\left(\underline{\hat{\mathbf{X}}}_r - E\left[\underline{\hat{\mathbf{X}}}_r\right]\right)^T \right] \right) = \frac{r}{N}\sigma_N^2$$

Este punto pretende analizar el impacto en la varianza como consecuencia de la proyección realizada sobre las observaciones. Para compactar la notación, definimos:

$$\underline{\boldsymbol{\beta}} = \underline{\hat{\mathbf{X}}}_r - E\left[\underline{\hat{\mathbf{X}}}_r\right] = \underline{\underline{\mathbf{P}}}_r\underline{\hat{\mathbf{X}}} - \underline{\underline{\mathbf{P}}}_r\underline{\mathbf{X}} = \underline{\underline{\mathbf{P}}}_r\underline{\mathbf{N}} \Rightarrow tr\left\{\underline{\boldsymbol{\beta}}\underline{\boldsymbol{\beta}}^T\right\} = \sigma_n^2 tr\left\{\underline{\underline{\mathbf{P}}}_r\underline{\underline{\mathbf{P}}}_r^T\right\} = \sigma_n^2 tr\left\{\underline{\underline{\mathbf{P}}}_r^T\underline{\underline{\mathbf{P}}}_r\right\} =$$

$$= \sigma_n^2 tr\left\{\underline{\underline{\mathbf{I}}}_{rxr}\right\} = r\sigma_n^2$$

y:

$$\sigma_r^2 = \frac{1}{N} tr\left\{\underline{\boldsymbol{\beta}}\underline{\boldsymbol{\beta}}^T\right\} = \frac{r}{N}\sigma_n^2$$

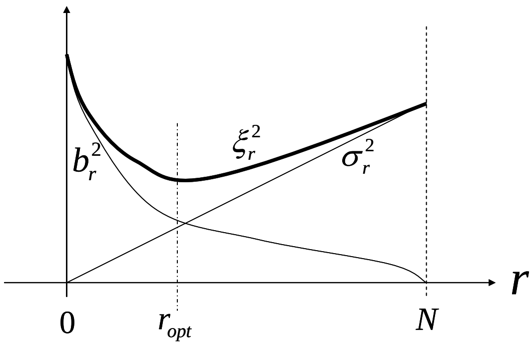
El resultado es el contrario al observado en (5.). La varianza aumenta con 'r'. La interpretación es que los modos asociados a los autovalores menores están más degradados por ruido. Su supresión ayuda a mejorar la estimación en términos de varianza.

6. El *error cuadrático medio* cuando tomamos la proyección $\hat{\underline{\mathbf{X}}}_r = \underline{\underline{\mathbf{P}}}_r\left(\underline{\mathbf{X}}+\underline{\mathbf{N}}\right)$ como una estimación de $\underline{\mathbf{X}}$ vendrá dado por $\xi_r^2 = b_r^2 + \sigma_r^2$. Esboce la evolución del *sesgo* y de la *varianza* en un gráfico en función del valor de r en el intervalo $0 \le r \le N$. Justifique que el *error cuadrático medio* presenta un mínimo para un posible valor de r.

Cuando consideramos tanto el sesgo como la varianza, tenemos que el e.c.m. viene dado por:

$$\xi_r^2 = b_r^2 + \sigma_r^2 = \frac{1}{N}\left(\sum_{i=r+1}^{N}\lambda_i + r\sigma_n^2\right)$$

Dado que los autovalores son no negativos, cualitativamente:



que ilustra la existencia de un óptimo en términos de e.c.m. en el que existe un compromiso entre sesgo y varianza, como para cualquier problema de estimación.

7. Si la relación *SNR* es arbitrariamente pequeña ($SNR \to 0$), justifique que sería razonable tomar $r = 0$, es decir, para $\hat{\underline{\mathbf{X}}}_0 = \underline{\mathbf{0}}$, o dicho de otro modo, la estimación $\hat{\underline{\mathbf{X}}}_0 = \underline{\mathbf{0}}$ es mejor que la propia observación $\hat{\underline{\mathbf{X}}} = \underline{\mathbf{X}} + \underline{\mathbf{N}}$ en un sentido de *mínimo error cuadrático medio*.

Veamos que ocurre cuando el término de ruido es dominante, es decir, $SNR \to 0$. Hemos visto en (3.):

$$SNR = \frac{tr\left(E\left[\underline{\mathbf{X}}\,\underline{\mathbf{X}}^T\right]\right)}{tr\left(E\left[\underline{\mathbf{N}}\,\underline{\mathbf{N}}^T\right]\right)} = \frac{1}{N\sigma_n^2}\sum_{i=1}^{N}\lambda_i \to 0 \Rightarrow \sigma_n^2 >> \frac{1}{N}\sum_{i=1}^{N}\lambda_i > \frac{1}{N}\sum_{i=r+1}^{N}\lambda_i$$

De (6.):

$$\xi_r^2 = b_r^2 + \sigma_r^2 = \frac{1}{N}\left(\sum_{i=r+1}^{N}\lambda_i + r\sigma_n^2\right) \approx \frac{1}{N}r\sigma_n^2 \;\; para \;\; SNR \to 0.$$

De donde concluimos que $\hat{\underline{\mathbf{X}}}_0 = \underline{\mathbf{0}}$, asociado a $r = 0$, tendería a ser la mejor estimación bajo un criterio de m.e.c.m.