Final Signals Brown, Alexander

1) Given:

a) $U = e^{jM}$ where M is Hermitian symmetric

b) - e^{A} where A is antihermitian
- $A^{H} = -A$

Find:

Show (a) and (b) are unitary matrices.

Solution:

a) A Hermitian symmetric matrix means that $M^{H} = M$

Theorem 6.2: Every $m \times m$ Hermitian symmetric matrix A can be diagonalized by a unitary matrix:

$$U^{H} A U = \Lambda$$

where U is unitary and Λ is diagonal.

Furthermore

$$A = U \Lambda U^{H} = \sum_{i=1}^{m} \lambda_i u_i u_i^{H}$$

Thus we can write

$$U = \sum_{i=1}^{n} e^{j\lambda_i} a_i a_i^{H}$$ where $a_i \in A$ and A is unitary

We want to show U is unitary, that means U is unitary if $UU^{H} = I$.

$\Rightarrow \quad U^H = \sum_{i=1}^{n} \left(e^{j\lambda_i} a_i a^H \right)^H = \sum_{i=1}^{n} e^{-j\lambda_i} a_i^H a_i$

$$= e^{-jH}$$

$\Rightarrow \quad UU^H = e^{jH} e^{-jH} = e^{jH - jH} = e^0 = I$

Thus U is unitary

b) Very similarly to (a) we have

$$A^H = -A \quad \text{and} \quad e^A$$

Let $U = e^A$

$UU^H = e^A e^{-A} = e^0 = I$

Thus e^A is unitary for antihermitian matrices and $A^H = -A$.

2) Given:

- X is a vector space with real variable t and real coefficients of degree one or less, that is an element $x(t) \in X$ is of the form

$$x_0 + x_1 t$$

- $Y = \mathbb{R}^2$
- Define a linear transformation $L: X \to Y$ by $L(c_0 + c_1 t) = [c_0 + c_1,\ c_0 - c_1]^T$
- Define the inner product as

$$\langle x(t), y(t) \rangle_X = \int_0^1 x(t) y(t)\, dt$$

- Define the inner product associated with the Y space as

$$\langle x, y \rangle_Y = x_0 y_0 + x_1 y_1$$

Find:

For these vector spaces, linear operator L, and inner products explicitely determine

- The adjoint L^*
- $R(L)$ and $R(L^*)$
- Justify your answer

Solution:

Let

$$x(t) = [x_0 \quad x_1] \begin{bmatrix} 1 \\ t \end{bmatrix}$$

be written as $[x_0 \quad x_1]^T = \underline{x}$

Definition 4.4: Let $A: X \to Y$ be a bounded linear operator from the Hilbert space X to the Hilbert space Y. The adjoint operator A, denoted *, is the operator $A^*: Y \to X$ such that

$$\langle Ax, y \rangle = \langle x, A^* y \rangle$$

for all $x \in X$ and $y \in Y$. An operator is self adjoint if $A^* = A$.

$$L\underline{x} = \begin{bmatrix} x_0 + x_1 \\ x_0 - x_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \Rightarrow L^* = L^H = L$$

page 238

The adjoint of a matrix is the conjugate transpose

page 244

$R(L)$ is the column space of L
$R(L^H)$ is the row space of L

$R(L^H) = \text{range}\{0\}$ because the vectors in L are orthogonal

$$R(L) = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$$

3) Given:

- A is real $n \times n$ symmetric positive definite matrix
- $A = U \Lambda U^T$ is the eigendecomposition
- Λ is ordered from largest to smallest
- $U = [u_1 \; u_2 \cdots u_n]$
- Find optimum rank-1 approximation to A which is denoted bb^T.
- Choose to minimize

$$\min_b \| A - bb^T \|_F^2$$

If it helps, let $b = \gamma^{1/2} c$ where c is unit length. In this case, formulate

$$\min_b \| A - \gamma cc^T \|_F^2 \quad \text{subject to } c^T c = 1$$

Find:

Use techniques of vector calculus to find optimum value of b.

Solution:

Let $A_1 = bb^T \Rightarrow \min_{A_K} \| A - A_1 \|_F^2$

$$\Rightarrow \| A - A_1 \|_F^2 = \| U \Lambda U^T - A_1 \|_F^2 = \| \Lambda U^T \; U^T A_1 \|_F^2$$

$$= \| \Lambda - U^T A_1 U \|_F^2 = \| \Lambda \|_F^2 - 2\gamma \Lambda^T cc^T + \gamma^2 \| cc^T \|_F^2$$

where $\| cc^T \|_F^2 = 1$ because they are unit vectors

Furthermore,

$$\Lambda cc^T = \sum_{i=1}^{r} \lambda_i c_i c_i \leq \lambda_i \sum_{i=1}^{r} |c_i||c_i| = \lambda_i \langle c, c \rangle$$

Thus by Cauchy-Schwarz

$$\langle c, c \rangle \leq \|c\| \, \|c\| = 1$$

$$\Rightarrow \lambda_i \langle c, c \rangle = \lambda_i$$

Recall that in Λ, λ_i is ordered from largest to smallest $\therefore$ we use λ_1.

Therefore we can write

$$\Lambda cc^T \leq \lambda_1$$

Thus we can rewrite the expanded norm as

$$\|\Lambda - U^T A_1 U\|_F^2 = \|\Lambda\|_F^2 - 2\gamma \Lambda cc^T + \gamma^2 \|cc^T\|_F^2$$

$$\geq \|\Lambda\|_F^2 - 2\gamma\lambda_1$$

Which we can view as

$$\|\Lambda\|_F^2 - 2\gamma\lambda_1 + 0$$

Lets attempt to complete the square by writing

$$\|\Lambda\|_F^2 - (\gamma - \lambda_1)^2 - \lambda_1^2$$

As shown in appendix B. This means we reach a critical point at $\gamma = \lambda_1$. That means

$$\| \Lambda - U^T A_1 U \|_F^2 \geq \| \Lambda \|_F^2 - \lambda_1^2$$

The minimum is then at

$$\gamma = \lambda_1 \; ; \; c = e_1$$

$$\Rightarrow A_1 = bb^T = \gamma (Uc)(Uc)^T$$

$$= \lambda_1 u_1 u_1^T \quad //$$

Another Way

$\|A - bb^T\|_F^2 = \langle A - bb^T, A - bb^T \rangle_F$

$= tr(A^TA - A^Tbb^T - b^TbA + b^Tbbb^T)$

$= tr(A^TA) + tr(-A^Tbb^T - b^TbA) + tr(b^Tbbb^T)$

$= tr(A^TA) - tr(2bb^TA) + tr(b^Tbbb^T)$

$\frac{\partial}{\partial b}\left[tr(A^TA) - tr(2bb^TA) + tr(b^Tbbb^T)\right]$

$= 0 - 2b^TA + bb^Tb + b^Tbb^T$

$= -2b^TA + 2b^Tbb^T = 2b^T(-A + bb^T) = 0$

$A = bb^T \triangleq U\Lambda U^T = \sum \lambda_i u_i u_i^T$

$bb^T \approx \lambda_i u_i u_i^T$ //

4) Given:

- A is complex $m \times n$ matrix
- $b \neq 0$ is complex $m \times 1$ vector
- $m > n+1$
- $\text{rank}(A) = n$
- $b \notin R(A)$ (column space of A)
- Let $A = \sum_{i=1}^{n} \sigma_i u_i v_i^H$
- Let $[A|b] = \sum_{i=1}^{(n+1)} \bar{\sigma}_i \bar{u}_i \bar{v}_i^H$ (the bars mean they are different that their non-barred friends).
- l_2 norms are assumed unless specified

Find:

a) Under what conditions on α does $A^H A - \alpha I$ guarantee a positive definite matrix? Express answers in terms of singular values of A.

b) Solve the standard least-squares problem for the $n \times 1$ vector x.

$$\min_x \|b - Ax\|^2$$

Derive the normal equations for the problem of the form

$$Mx = c \quad \text{and} \quad \hat{x}_{ls}$$

c) Solve the modified LS problem

$$\min_x -\rho\|x\|^2 + \|b - Ax\|_2^2 \qquad \text{where } \rho \in \mathbb{R} \tag{1}$$

- Find normal equations in form

$$Mx = c \quad \text{and} \quad x_{mls}$$

d) We will now show that total least squares solution is obtained from (1) with the appropriate ρ.
We know the TLS solution

$$\begin{bmatrix} \hat{x}_{tls} \\ -1 \end{bmatrix} = \frac{-1}{\bar{v}_{n+1,n+1}} \bar{v}_{n+1}$$

- Since $[\hat{x}_{tls} \ -1]^T$ is a scaled version of a right singular vector, it is an eigenvector of what matrix?

- What is the corresponding eigenvalue?
- Write the appropriate eigenequation to show that $\hat{x}_{TLS}$ can be computed directly by

$$\hat{x}_{TLS} = (A^H A - \bar{\sigma}_{n+1}^2 I)^{-1} A^H b$$

e) Explain why TLS can be calculated by the optimization problem in (1). What value of ρ should be used?

Solution:

2) Theorem 7.2: Let A be an $m \times n$ matrix with rank$(A) = r$ and let $A = U \Sigma V^H$. Let $k < r$ and let

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^H = U \Sigma_k V^H$$

where

$$\Sigma_k = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k)$$

Then $\|A - A_k\| = \sigma_{k+1}$ and A_k is the nearest matrix of rank k to A.

$$\min_{\text{rank}(B)=k} \|A - B\|_2 = \|A - A_k\|_2$$

3/6

$A^H A - \alpha I \geq 0$

$A = U \Sigma V^H = \sum_{i=1}^{n} \sigma_i u_i v_i^H$

$A^H A = V \Sigma^H \underbrace{U^H U}_{\substack{I \text{ because} \\ U \text{ is unitary}}} \Sigma V^H = V \Sigma^H \Sigma V^H = V \Sigma^2 V^H$

$= \sum_{i=1}^{n} \sigma_i^2 v v^H \geq \sigma_n^2$

$A^H A = \alpha I \geq \sigma_n^2$

$\therefore \; A^H A - \alpha I \geq 0$ if $\alpha = \sigma_n^2$

b) First rewrite Ax

$Ax = \sum_{i=1}^{m} x_i p_i$ where $A = [p_1 \; p_2, \ldots, \; p_m]$

And recall $\|\cdot\|^2 = \langle \cdot , \cdot \rangle$ is the induced norm.
We write the problem as

$\min_x \| \underbrace{b - \sum_{i=1}^{m} x_i p_i}_{e} \|^2$

The error is desired to be orthogonal to the data $\therefore$

$\langle b - \sum_{i=1}^{m} x_i p_i , x_j \rangle = 0$

$\Rightarrow \langle \sum_{i=1}^{m} x_i p_i , x_j \rangle = \langle b, x_j \rangle$

Expanding the previous result we find

$$\underbrace{\begin{bmatrix} \langle p_1, p_1\rangle & \langle p_2, p_1\rangle & \cdots & \langle p_m, p_1\rangle \\ \langle p_1, p_2\rangle & \langle p_2, p_2\rangle & & \langle p_m, p_2\rangle \\ \vdots & & & \vdots \\ \langle p_1, p_m\rangle & \langle p_2, p_m\rangle & \cdots & \langle p_m, p_m\rangle \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} \langle b, x_1\rangle \\ \langle b, x_2\rangle \\ \vdots \\ \langle b, x_m\rangle \end{bmatrix}}_{c}$$

From which we recognize

$$M = A^H A \quad \text{and} \quad c = A^H b$$

$$\Rightarrow \quad A^H A x = A^H b \Rightarrow \hat{x}_{LS} = (A^H A)^{-1} A^H b$$

c) $\min_x -\rho \|x\|^2 + \|b - Ax\|^2$

Begin by rewriting in terms of the induced norm

$-\rho \langle x, x\rangle + \langle b - Ax, b - Ax\rangle$

$-\rho x^H x + (b - Ax)^H (b - Ax)$

$-\rho x^T x + x^T A^H A x - 2A^H b x - b^H b = f(x)$

Take the gradient of f with respect to x

$$\frac{\partial f}{\partial x} = -2\rho x + 2A^H A x - 2A^H b$$

5/6

Equate $\frac{\partial f}{\partial \underset{\sim}{x}} = 0$ to find the normal equation

$$-2\rho x + 2A^H A x - 2A^H b = 0$$

$$\underbrace{(A^H A - \rho)}_{M} x = \underbrace{A^H b}_{c}$$

$$\Rightarrow \quad \hat{x}_{MLS} = (A^H A - \rho)^{-1} A^H b$$

d) • $\bar{v}_{n+1}$ is the right singular vector of $[A\ b]$
• $[\hat{x}_{TLS}\ -1]^T$ is a scaled version of a right singular vector

$$\begin{bmatrix} \hat{x}_{TLS} \\ -1 \end{bmatrix} = \frac{-1}{v_{n+1,n+1}} \bar{v}_{n+1}$$

$\therefore$ $[\hat{x}_{TLS}\ -1]$ is a scaled singular vector of $[A\ b]$ and has an eigenvalue of $\bar{\sigma}_{n+1}$.

We can then write

$$([A\ b] - \bar{\sigma}_{n+1}) \begin{bmatrix} \hat{x}_{TLS} \\ -1 \end{bmatrix} = 0$$

$$\Rightarrow \quad [A\ b] \begin{bmatrix} \hat{x}_{TLS} \\ -1 \end{bmatrix} = \bar{\sigma}_{n+1} \begin{bmatrix} \hat{x}_{TLS} \\ -1 \end{bmatrix}$$ ~ The eigenvalue is $\bar{\sigma}_{n+1}$

$$\Rightarrow \quad \frac{1}{\bar{\sigma}_{n+1}} [A\ b] \begin{bmatrix} x_{TLS} \\ -1 \end{bmatrix} = \begin{bmatrix} \hat{x}_{TLS} \\ -1 \end{bmatrix} = \frac{-1}{v_{n+1,n+1}} \bar{v}_{n+1}$$

6/6

e) Much like the the ideology given in the book, the solution

$$\begin{bmatrix} \hat{x}_{TLS} & -1 \end{bmatrix}^T$$

must lie in the nullspace of the "C + D" thus the "D" must be perturbed enough such that the system is rank deficient. By finding the largest σ that keeps

$$A^H A - \sigma I \geq 0$$

it also perturbs the system enough to be rank deficient by removing all the "noise" degrees of freedom, or the degrees of freedom that have the least impact on minimizing the squared distance.

The value of ρ to be used is

$$\rho = \bar{\sigma}_{n+1}^2 \quad //$$

5) Given:

- A is an $m \times n$ full rank matrix
- $A = A(\phi)$
- A signal $x_t = A(\phi)s_t + n_t$
- $E[s_t s_t^H] = \text{diag}(p_1, p_2, \dots, p_n) = P$
- $n_t \sim N(0, \sigma_n^2)$

Find:

a) Show $R_x = E[x_t x_t^H] = APA^H + \sigma_n^2 I$

b) - Let $R_x = U \Lambda U^H$ be an eigendecomposition
- Λ has a non-decreasing order $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$
- Partition the eigenvalues and eigenvectors into 2 subsets
  - Signal subspace
  - Noise subspace
- How many eigenvalues are in each set?
- What are the noise eigenvalues equal to?

Solution:

a) We will write A instead of $A(\phi)$ for convenience

$$E[x_t x_t^H] = E[(As_t + n_t)(As_t + n_t)^H]$$

$$= E[(As_t + n_t)(s_t^H A^H + n_t^H)]$$

$$= E[As_t s_t^H A^H + As_t n_t^H + n_t s_t^H A^H + n_t n_t^H]$$

$$= E[As_t s_t^H A^H] + \underbrace{E[2As_t n_t^H]}_{\substack{\text{Expected value of the} \\ \text{cross terms are 0} \\ \text{because } n_t \text{ is white}}} + E[n_t n_t^H]$$

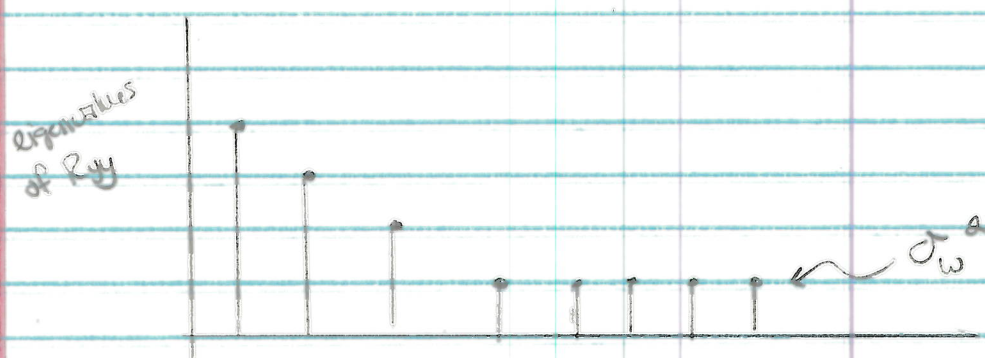
$= E[A s_t s_t^H A^H] + E[n_t n_t^H]$

$= APA^H + \sigma_n^2 I$ //

b) $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$

Let s denote the number of modes (eigenvalues) in the signal space Λ with $s < m$. Then there are $m-s$ eigenvalues in the noise space.

Because of the noise, the eigenvalues of the noise will be $\lambda_n = \sigma_n^2$ for $s < n \leq m$.
Much like the plot in class - 4-14-23



Where $r_{yy}[k] = r_{xx}[k] + \sigma_w^2 \delta[k]$

c) Let U_s be the matrix with columns formed by the signal eigenvectors. Show $R(U_s) = R(A(\phi))$

That is we want to show the range of the signal space is the same as the eigenvector space.

We know U_s is built on eigenvectors, so the matrix must be orthogonal. That is

$$A^T = A^{-1}$$

Begin by writing the definition of eigenvalues and eigenvectors

$$Ax = \lambda x \qquad (1)$$

where x denotes the eigenvector and λ the eigenvalue. To make this more general write (1) in terms of multiple eigenvalus and eigenvectors

$$Ax_i = \lambda x_i \qquad \forall\, i \in \text{rank}(A)$$

Let $U = [x_1\ x_2, \ldots, x_m] = [U_s\ U_N]$
$\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$

Substituting we find a form similar to the eigendecomposition

$$AU = U\Lambda$$

$$\Rightarrow\ A = U\Lambda U^{-1} = U\Lambda U^{+1}$$

Which also looks a lot like the SVD. By (7.7) in the book

$$R(A) = \text{span}(U_1)$$
$$N(A) = \text{span}(V_2)$$
$$R(A^{+1}) = \text{Span}(V_1)$$
$$N(A^{+1}) = \text{span}(U_2)$$

Lets write down what is meant by U_1, U_2, V_1, V_2

For a general matrix B $\quad rank(B) = p \quad p < \min(m,n)$

$$\Sigma = diag(\sigma_1, \sigma_2, \ldots, \sigma_p, 0, \ldots, 0)$$

$$B = U \Sigma V^H = [U_1 \; U_2] \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^H \\ V_2^H \end{bmatrix}$$

Where we can see span(U_1) is the range of the signal space. In our case $U_1 = U_s$ $\therefore$

$$R(A) = R(U_s) \quad //$$

6) Given:

- x is a 2x1 random vector
- $E[x] = 0$
- $R = E[xx^T]$
- R has the eigendecomposition

$$R = U \Lambda U^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \frac{1}{\sqrt{2}}$$

- Let x be corrupted by noise

$$y = x + v$$

where v is a random vector with

- $E[v] = 0$; $E[vv^T] = \sigma_v^2 I$

- Take an observed vector y and process it to obtain a scale measurement using a vector w

$$z = w^T y = w^T x + w^T v = z_{signal} + z_{noise}$$

- We wish to pick the vector w to maximize the signal strength to noise ratio. That is

$$SNR = \frac{E[z^2_{signal}]}{E[z^2_{noise}]}$$

Find:

- What is the unit-length w that will maximize the SNR?
- What is the corresponding max SNR?

Solution:

$$SNR = \frac{E[z_s^2]}{E[z_n^2]} = \frac{E[z_s z_s^T]}{E[z_n z_n^T]}$$

$$= \frac{E[W^T x x^T W]}{E[W^T v v^T W]}$$

Because we are "weighting" by w, it has no uncertainty so we can pull it out

$$\Rightarrow \frac{W^T E[x x^T] W}{W^T E[v v^T] W} = \frac{W^T R W}{W^T \sigma_n^2 W}$$

Note that this looks very similar the Rayleigh quotient which is of the form

$$R(x) = \frac{x^H A x}{x^H x}$$

with assistance from Theorem 6.5

Theorem: for a positive semi-definite self-adjoint matrix A with $Q_A(x) = \langle Ax, x \rangle = x^H A x$, the maximum

$$\max_{\|x\|_2 = 1} Q_A(x)$$

is λ_1, the largest eigenvalue of A, and the maximizing x is $x = x_1$, the eigenvector corresponding to λ_1.

Furthermore if we maximize $Q_A(x)$ subject to the constraints

1) $\langle x, x_j \rangle = 0 \quad j = 1, 2, \ldots, k-1$ and
2) $\|x\|_2 = 1$

then λ_k is the maximized value subject to the constraints and x_k is the corresponding value of x.

That is from the Rayleigh quotient we find

$$\max_{\|x\| \neq 0} R(x) = \lambda_1 \quad \text{and the maximizing } x^* = x_1.$$

Rewriting the last result from our problem we see that we have the Rayleigh quotient

$$SNR = \frac{W^T R W}{\sigma_n^2 W^T W}$$

Thus the max SNR is

$$SNR = \frac{\lambda_1}{\sigma_n^2} \quad \text{with } x^* = x_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$\lambda_1 = 3$

7) Given:

- u_i and u_i' represent the same point on different coordinates
- This problem uses projective coordinates

$$\lambda(u, v, 1) \text{ represents } (u, v) \quad \forall \lambda \neq 0$$

- The homography between the points is defined as

$$\alpha_i u_i' = H u_i \qquad i = 1, 2, \ldots, m$$

α_i is a scale parameter that adds a degree of freedom into the computations to allow for scale factors

- $G(u')\alpha u' = \alpha G(u')u' = 0 = G(u')Hu$

$$G(u') = \begin{bmatrix} 0 & -w & v \\ w & 0 & -w \\ -v & u & 0 \end{bmatrix} \quad \text{for } u = [u\ v\ w]^T \qquad (1)$$

- Chapter 9 stuff says we can write

$$\underbrace{\begin{bmatrix} (u_1^T \otimes G(u_1')) \\ (u_2^T \otimes G(u_2')) \\ \vdots \\ (u_m^T \otimes G(u_m')) \end{bmatrix}}_{W} H = 0 \qquad (2)$$

- We are finding a nonzero H that lies in the null space of W.

Find:

a) Show $G(u)$ in (1) satisfies $G(u)u = 0$

b) Show how to use the SVD to solve

minimize $\|WH\|^2$ s.t. $\|H\|^2 = 1$

Solution:

a)

$$\begin{bmatrix} 0 & -w & v \\ w & 0 & -u \\ -v & u & 0 \end{bmatrix}\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 - wv + wv \\ wu + 0 - wu \\ -vu + vu \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

b) $\min \|WH\|_2^2$ s.t. $\|H\|_2^2 = 1$

Some facts about SVD

$$A = U\Sigma V^H = \sum_{i=1}^{r} \sigma_i u_i v_i^H \qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$$

$v_{k+1} \perp v_k$

This problem is very resemblant of the rank reduction problem presented in 7.5 in the book. We have a full rank matrix corrupted by noise. We know that the signal space is orthogonal to the Noise space. Thus we write

$$WH = \sum_{i=1}^{k+1} \sigma_i (v_i^H H) u_i$$

From 7.5 we see

$$\|A - B\|_2^2 \geq \|AH\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 (v_i z)^2 \geq \sigma_{k+1}^2$$

where $z \in \operatorname{span}(v_1, v_2, \ldots)$

Rewriting in terms of our problem. Let $\|h\|_2 = 1$

$$\|Wh\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 (v_i^H h)^2 \geq \sigma_{k+1}^2$$

where $h \in \text{span}(v_1, v_2, \ldots)$.

That is $\sigma_{k+1} \ll \sigma_k$ meaning that the first k σ values have the largest contribution to the solution. Using the corresponding vector $h = v_{k+1}$ (noting that its orthogonal to all other v's other than itself) results in

$$Wh = 0 = Wv_{k+1}$$

With the minimum norm.

Another way

$$\mathcal{L} = h^H W^H W h - \lambda h^H h \rightarrow \frac{\partial \mathcal{L}}{\partial h^H} = W^H W h = \lambda h$$

Where we recognize that this resembles the eigen equation where h is the eigenvector. Pick v_{k+1} to minimize.