Numerical Solution to the Inverse Problem:

Spectral Expansion and TSVD

Numerical Solution to the Inverse Problem.

Goal: minimize $\quad \phi = \phi_m + \beta \phi_d$

such that $\quad \phi_d = \phi_d^*$

where β is a tradeoff parameter $\beta > 0$, ϕ_d^* is a target misfit.

$$\phi = \| W_d (Gm - d^{obs}) \|^2 + \beta \| W_m (m - m_0) \|^2 \qquad (2)$$

where W_d is $N \times N$ matrix

W_m is an $L \times M$ matrix. (L can be: $L < M$, $L = M$, $L > M$)

Taking the gradient of (2) and setting the result equal to zero

$$\nabla_m \phi = \cancel{2} G^T W_d^T W_d (Gm - d^{obs}) + \cancel{2} \beta W_m^T W_m (m - m_0) =$$

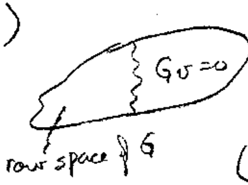
or
$$\boxed{\left(G^T W_d^T W_d G + \beta W_M^T W_m \right) m = G^T W_d^T W_d d^{obs} + \beta W_m^T W_m m_0} \qquad (3)$$

The l.h.s is an $M \times M$ symmetric and (almost always) positive definite matrix for non-zero β. It is positive definite if any of the following hold

(i) W_m^{-1} exists. (W_m is $M \times M$)

(ii) W_m has rank M (W_m is $L \times M$, $L > M$)

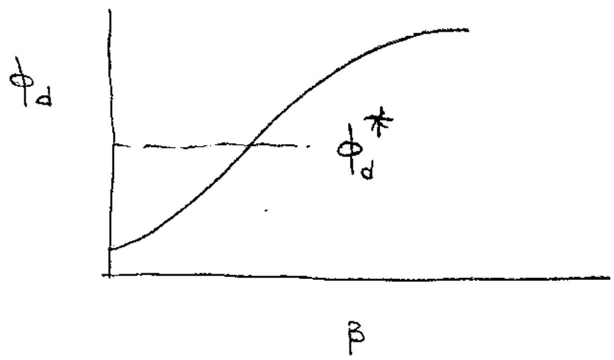
(iii) $N(W_m) \in$ Row space $(W_d G)$

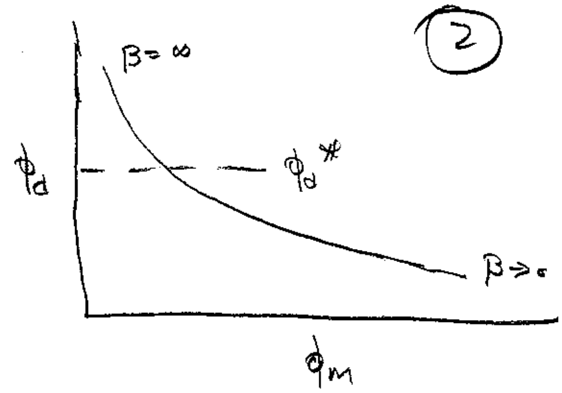


So the matrix can be inverted and

$$\boxed{m = \left(G^T W_d^T W_d G + \beta W_m^T W_m \right)^{-1} \left(G^T W_d^T W_d d^{obs} + \beta W_m^T W_m m_0 \right)} \qquad (4)$$

As discussed earlier when we introduced Tikhonov regularization, the system (3) is solved for different values of β and a line search is carried out to find β^* such that $\phi_d = \phi_d^*$.





Remark: For small problems ($M \sim 100$'s) the computation can be sufficiently fast so that

For large problems however using this methodology might be computationally prohibitive. Moreover there is a numerical worsening of the condition number of the matrix because we are working with G^TG rather than the matrix G. This leads us to work with an SVD solution to the problem.

Spectral Expansion Solution

The technique is motivated by the following reasoning. Given the equation

$$Gm = d$$

and an underdetermined problem (G: $N \times M$ matrix) we found a minimum norm solution (minimize $\phi_m = \|m\|^2$) using SVD

$$G = U \Lambda V^T$$

$$m_c = m^{\parallel} = VV^T m = V \Lambda^{-1} U^T d$$

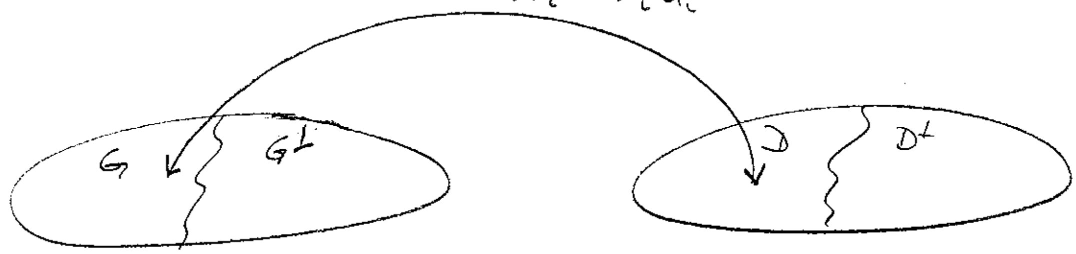
or

$$\boxed{m_c = \sum_{i=1}^{N} \left(\frac{u_i^T d}{\lambda_i} \right) v_i} \qquad N = \text{rank}\{G\}$$

The generalized inverse provided a model which fit the data exactly, provided that the data were consistent ($d \in Col(G)$)

We recognized that when λ_i becomes small that the coefficient $\left(\frac{u_i^T d}{\lambda_i}\right)$ can become large if d is contaminated with noise. This can cause unnecessarily large amounts of the basis vector $\vec{v}_i$ to be included.

$G\vec{v}_i = \lambda_i \vec{u}_i$

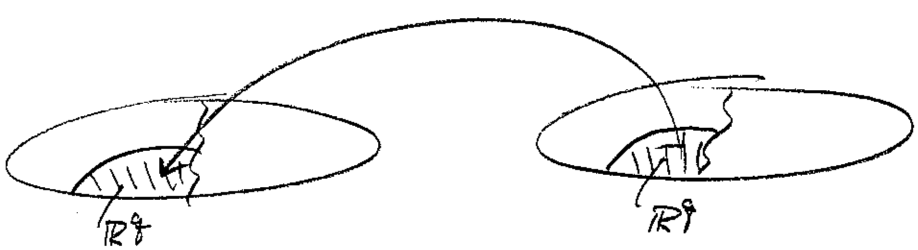


$$m_c = \sum_{i=1}^{q} \left(\frac{u_i^T d}{\lambda_i}\right) v_i + \underbrace{\sum_{i=q+1}^{N} \left(\frac{u_i^T d}{\lambda_i}\right) v_i}$$

because of the small λ_i, these vectors effectively produce more harm than they do good!

The spectral expansion thus invites an approximate solutions, by simply winnowing the contributions associated with singular values that are too small.

Thus $$m_c = \sum_{i=1}^{q} \left(\frac{u_i^T d}{\lambda_i}\right) v_i$$



Effectively we reduce the problem by working only in a q-dimensional subspace of the original spaces. This will called a Truncated SVD (TSVD) solution.

SVD Solution & Reduction to Standard Form.

$$\boxed{\begin{array}{lll} \text{Minimize} & \phi_m = \| W_m (m - m_0) \|^2 & \\ \text{subject to} & \phi_d \equiv \| W_d (Gm - d^{obs}) \|^2 = \phi_d^* \end{array}} \qquad (1)$$

Make a series of transformations.

Let $\boxed{x = W_m (m - m_0)}$

so $\boxed{m = W_m^{-1} x + m_0}$ Note that this requires that W_m be invertible but this poses no difficulty.

$$W_d (Gm - d^{obs}) = W_d \left(G [W_m^{-1} x + m_0] - d^{obs} \right)$$

$$= W_d G W_m^{-1} x + W_d G m_0 - W_d d^{obs}$$

Define $\boxed{\begin{array}{l} A \equiv W_d G W_m^{-1} \\ b = -W_d G m_0 + W_d d^{obs} \end{array}}$

With these definitions the optimization problem in (1) becomes

$$\text{minimize} \quad \phi = \| x \|^2$$

$$\text{subject to} \quad \| Ax - b \|^2 = \phi_d^*$$

or

$$\text{minimize} \quad \phi = \| Ax - b \|^2 + \beta \| x \|^2$$

take ∇_{x} $\quad \cancel{2} A^T (Ax - b) + \cancel{2} \beta x = 0$

or $\boxed{(A^T A + \beta I) x = A^T b}$

Introduce the SVD of A: $A = U\Lambda V^T$

$$A^TA = V\Lambda U^TU\Lambda V^T = V\Lambda^2 V^T$$

so $$\left(V\Lambda^2 V^T + \beta I\right)x = V\Lambda U^T b$$

$\times V^T$ $$\left(\Lambda^2 V^T + \beta I V^T\right)x = \Lambda U^T b \qquad (V^TV = I_p)$$

$$\left(\Lambda^2 + \beta I\right)V^T x = \Lambda U^T b$$

$$V^T x = \left(\Lambda^2 + \beta I\right)^{-1}\Lambda U^T b \qquad (1a)$$

Solution becomes $$\boxed{V V^T x = x_R = V\left(\Lambda^2 + \beta I\right)^{-1}\Lambda U^T b} \qquad (2)$$

($x_R \equiv x$ regularized)

Notice the difference between (2) and the SVD solution of

$$Ax = b$$
$$U\Lambda V^T x = b$$
$$\boxed{x_{SVD} = V\Lambda^{-1}U^T x} \qquad (3)$$

The difference lies only in the value of the diagonal matrix. We can obtain the Tikhonov regularized solution x_R by modifying equation 3.

$$x_R = V\left(\Lambda^2 + \beta I\right)^{-1}\Lambda U^T b \doteq V\left\{(\Lambda^2 + \beta I)^{-1}\Lambda^2\right\}\Lambda^{-1}U^T b$$

$$\boxed{x_R = V T \Lambda^{-1} U^T b}$$

where $T = \text{diag}(t_1, t_2, \ldots t_p)$. where $0 \le t_i \le 1$

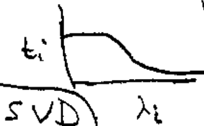
and for the Tikhonov regularization considered here

$$\boxed{t_i = \frac{\lambda_i^2}{\lambda_i^2 + \beta}}$$

The values t_i are often referred to as the filter parameters. The general regularized solution can be written as

$$\boxed{x_c = \sum_{i=1}^{N} \frac{t_i}{\lambda_i} (u_i^T b)\, v_i}$$

$t_i \approx 1$ for $\lambda_i^2 \gg \beta$
$t_i \approx 0$ for $\lambda_i^2 \ll \beta$



So $t_i = 0 \Rightarrow$ no contribution from the vector v_i
$\quad = 1 \Rightarrow$ full contribution as determined from SVD

<u>Misfit for underdetermined systems.</u> (and compatible equations)

For this situation, all of data space is activated. $p = n$ and $UU^T = I$. The constructed solution is

$$x_c = V T \Lambda^{-1} U^T b$$

Let the predicted data be

$$\begin{aligned} b_c &= A x_c \\ &= U \Lambda V^T V T \Lambda^{-1} U^T b \\ &= U \Lambda T \Lambda^{-1} U^T b \end{aligned}$$

$$\boxed{b_c = U T U^T b}$$

The misfit is

$$\begin{aligned} \phi_d = \| b - b_c \|^2 &= (b - b_c)^T (b - b_c) \\ &= (b - b_c)^T U U^T (b - b_c) \\ &= [U^T (b - b_c)]^T [U^T (b - b_c)] \\ &= \| U^T (b - b_c) \|^2 \\ &= \| \hat{b} - \hat{b}_c \|^2 \end{aligned}$$

$\underline{\underline{UU^T = I}}$

So $\boxed{\phi_d = \| \hat{b} - \hat{b}_c \|^2}$

$\{$ or $\| b \| = \| U^T b \|$ for U orthogonal $\}$

We had $$b_c = U T U^T b$$
$$U^T b_c = T U^T b$$

so $$\boxed{\hat{b}_c = T \hat{b}}$$

So $\phi_d = \| b - b_c \|^2 = \| \hat{b} - \hat{b}_c \|^2 = \| \hat{b} - T\hat{b} \|^2$

or $$\boxed{\phi_d = \sum_{i=1}^{N} (1-t_i)^2 \hat{b}_i^2}$$

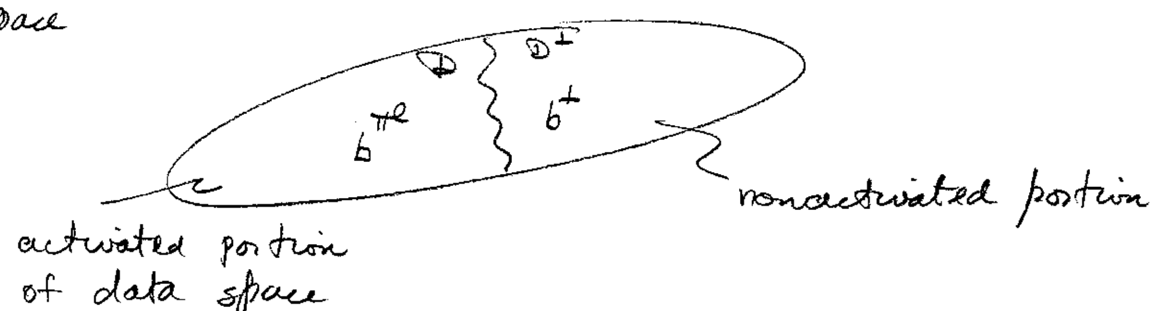
For the choice of t_i used here we have

$$1 - t_i = 1 - \frac{\lambda_i^2}{\lambda_i^2 + \beta} = \frac{\lambda_i^2 + \beta - \lambda_i^2}{\lambda_i^2 + \beta} = \frac{\beta}{\beta + \lambda_i^2}$$

$$\boxed{\phi_d = \sum_{i=1}^{N} \left(\frac{\beta}{\beta + \lambda_i^2} \right)^2 \hat{b}_i^2}$$

Misfit for overdetermined systems or imposed maximum rank

Data space



We saw that the projection of b onto the activated portion of data space is
$$b^{\parallel} = U U^T b$$

Thus $$\boxed{b^\perp = (I_N - U U^T) b}$$

This component of the data can never be fit. It constitutes a minimum error

$$\boxed{\phi_d^{min} = \| b^{\perp} \|^2 = \| (I_N - UU^T) b \|^2}$$

As such, for a matrix system of useable rank p where $p \le M$ and $p \le N$ the misfit from an SVD solution is

$$\boxed{\phi_d = \sum_{i=1}^{p} (1 - t_i)^2 \hat{b}_i^2 + \| (I_N - UU^T) b \|^2}$$

Model Norm

The last quantity to solve for is the model norm.

$$\phi_m = \| x \|^2 = \| W_m (m - m_o) \|^2$$

Our solution was $\quad x_c = V T \Lambda^{-1} U^T b = V T \Lambda^{-1} \hat{b}$.

$$\| x_c \|^2 = x_c^T x_c = \hat{b}^T \Lambda^{-1} T V^T V T \Lambda^{-1} \hat{b}$$
$$= \hat{b}^T T^2 \Lambda^{-2} \hat{b}$$

$$\boxed{\phi_m = \sum_{i=1}^{N} \left(\frac{t_i}{\lambda_i} \right)^2 \hat{b}_i^2}$$

Again, for our specific choice of $t_i = \dfrac{\lambda_i^2}{\lambda_i^2 + \beta}$ we get the norm for our minimization.

Nature of the SVD Solution and Truncated SVD.

Consider a general matrix G that results from discretizing a linear inverse problem.

$$d_j = \int g_j(x)\, m(x)\, dx \qquad j = 1, N$$

$$G \sim \begin{pmatrix} \text{---} \; g_1 \; \text{---} \\ \text{---} \; g_2 \; \text{---} \\ \vdots \\ \text{---} \; g_N \; \text{---} \end{pmatrix}$$

Rows of G are essentially the kernel f^{cns}. They tend to be smoothing type f^{cns}.

The SVD representation

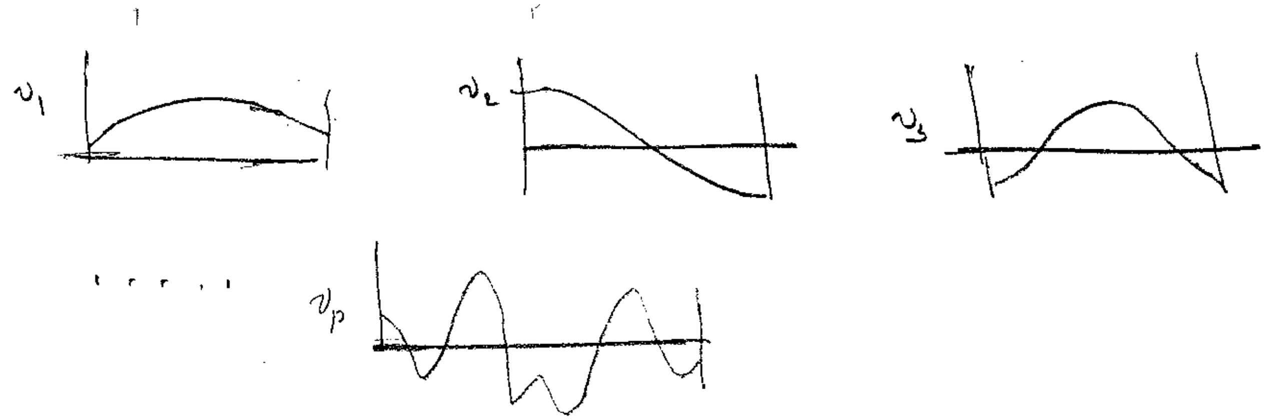
$$G = U_p \Lambda_p V_p^T$$

$$V = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_p \\ | & & | \end{pmatrix} \qquad \Lambda_p = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix} \qquad U_p = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_p \\ | & & | \end{pmatrix}$$

$(M \times p) \qquad (p \times p) \qquad N \times p$

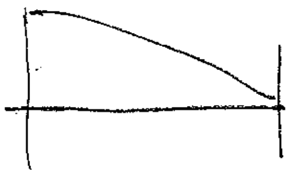
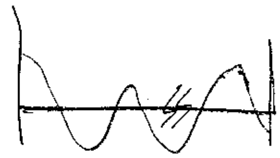
The ordering of the singular values is $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$

In general the eigenvectors $\{v_i\}$ also have an ordering



That is, the number of zero crossings increases with singular value number

The same character is observed for the vectors $\{u_i\}$

u_1  . . . . . . u_p 

Now look at the SVD solution for the regularized problem

$$x_c = V T \Lambda^{-1} U^T b \qquad\qquad b_c = A x_c = U T U^T b$$

$$x_c = \sum_{i=1}^{N} \frac{t_i}{\lambda_i} (u_i^T b)\, v_i$$

$$\phi_d = \sum_{i=1}^{N} (1 - t_i)^2 (u_i^T b)^2 \qquad\qquad b_c = \sum t_i (u_i^T b)\, u_i$$

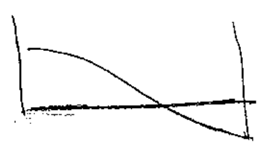
$$\phi_m = \sum_{i=1}^{N} \left(\frac{t_i}{\lambda_i}\right)^2 (u_i^T b)^2$$

With this decomposition we can see precisely what the effect of keeping each eigenvector v_i in the solution

Consider the i^{th} eigenvector v_i

$$\Delta x_c = \frac{t_i}{\lambda_i} (u_i^T b)\, v_i \qquad \Delta\phi_d = (1 - t_i)^2 (u_i^T b)^2 \qquad \Delta\phi_m = \left(\frac{t_i}{\lambda_i}\right)^2 (u_i^T b)$$

$$\Delta b_c = t_i (u_i^T b)\, u_i$$

for i small. v_i  has small number of zero crossings.

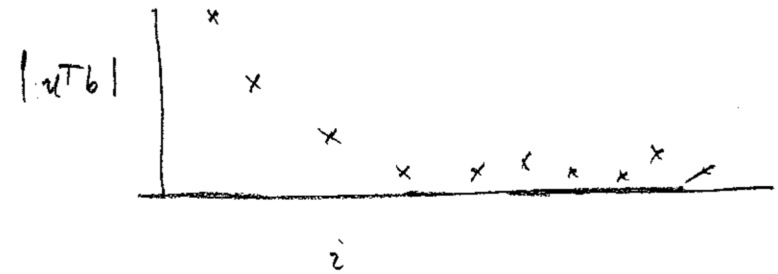
So, this is a "smooth" or large-scale structural addition to the model. <u>We like this vector</u>!

Moreover, the amount of this vector added to the solution is controlled

for i small $\Rightarrow$ λ_i is (relatively) large

So $\frac{(u_i^T b)}{\lambda_i}$ is not a large number.

Generally the data contain some large scale structure. What is generally perceived is



$|u_i^T b|$ is "large" for small index 'i'.

Thus the contribution

$$\Delta b_c \sim (u_i^T b) \quad \text{is large} \qquad (\text{we like this!})$$

but

$$\Delta \phi_m \sim \frac{(u_i^T b)^2}{\lambda_i^2} \quad \text{is small} \qquad (\text{we like this!})$$

So the contributions to the solution provided by the first few eigenvectors has many desirable properties

(a) adds a smooth large scale structural component

(b) greatly reduces the misfit $\qquad (\phi_d = \| b - b_c \|^2)$

(c) doesn't substantially increase the model norm.

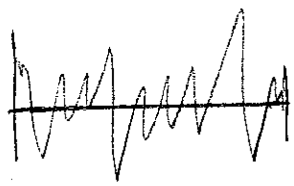
So for these eigenvectors we want $t_i \sim 1$

For Tikhonov regularization

$$t_i = \frac{\lambda_i^2}{\lambda_i^2 + \beta}$$

So $t_i \simeq 1$ if $\beta << \lambda_i^2$.

Now consider what happens for large 'i' (associated with small λ_i)

v_i  has lots of structure (we don't generally want this character in the model)

$\Delta b_c \sim (u_i^T b)$ is small because $|u_i^T b|$ is small.

$\Delta \phi_m \sim \dfrac{(v_i^T b)}{\lambda_i^2}$ is large because λ_i is small.

So the contibutions to the solution provided by the last few eigenvectors has many undesireable properties

(a) adds a high structural component

(b) doesn't significantly reduce the misfit

(c) greatly increases the model norm

So for these eigenvectors we want $t_i \sim 0$

For Tikhonov regularization

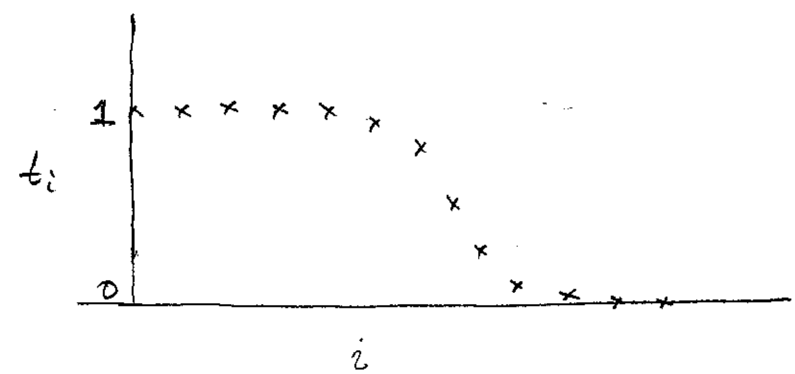
$$t_i = \frac{\lambda_i^2}{\lambda_i^2 + \beta}$$

So $t_i \simeq 0 \Rightarrow \beta \gg \lambda_i$

Truncated SVD.

The previous analysis shows that the filter factors should be unity for the first few singular vectors and should be close to zero for the last few. One could think of other regularization schemes that adhered to these principals and might accomplish the same general result as the Tikhonov filter coefficients

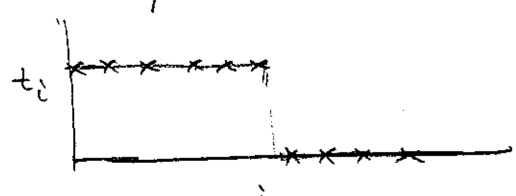
$$t_i = \frac{\lambda_i^2}{\lambda_i^2 + \beta}$$



Rather than have all components $\{v_i\}$ affected by the regularization (so the 1^{st} one is not incorporated with unit amplitude and there is still some vestage of the smallest eigenvectors) we could adopt a "keep or discard" strategy.

ie $\quad t_i = 1 \quad i = 1, p$

$\quad\;\; t_i = 0 \quad i > p$



The value of p (ie the number of basis vectors to keep) could be determined by evaluating the misfit.

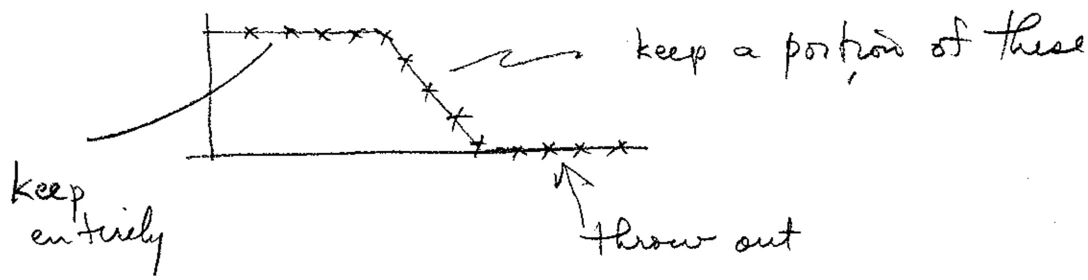
misfit $\quad \boxed{\phi_d = \sum_{i=p+1}^{N} (1-t_i)^2 (u_i^T b)^2}$

Find a value of p such that $\phi_d \simeq \phi_d^*$.

In this case the model is

$$\boxed{x_c = \sum_{i=1}^{p} \frac{(u_i^T b)}{\lambda_i} v_i}$$

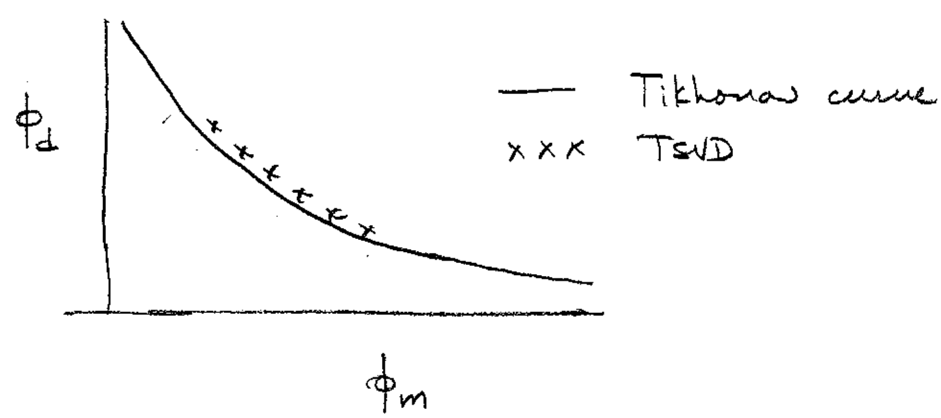
Remark: You could choose other 'ad hoc' methods also..



In the above scenario you could adjust the filter coefficients for basis vectors in the transition zone so the misfit $\phi_d = \phi_d^*$.

Note: By using TSVD it is unusual to find a 'p' that yields "exactly" the misfit $\phi_d = \phi_d^*$.

Remark: When working with TSVD, the plot of the tradeoff between ϕ_m and ϕ_d as a function of p is often observed to be quite similar to that obtained from Tikhonov regularization



Remark: The TSVD solutions (or any other ad hoc filtering solution) must always have a (ϕ_d, ϕ_m) coordinate that lies above the Tikhonov curve. The latter is optimal since it is an exact minimization of $\phi = \phi_d + \beta\phi_m$.

Least Squares Representation

Consider $\phi = \| Gm - d \|^2 + \beta \| W(m - m_0) \|^2$

minimizing ϕ yields $\boxed{(G^TG + \beta W^TW)m = G^Td + \beta W^TW m_0}$ (1)

to be solved where the matrix on the left is $M \times M$. The solution obtained via this route is identical to solving the over determined system

$$\boxed{\begin{bmatrix} G \\ \sqrt{\beta}\, W \end{bmatrix} m = \begin{bmatrix} d \\ \sqrt{\beta}\, W m_0 \end{bmatrix}} \qquad (2)$$

To see this, the least squares misfit solution to $Ax = b$ $A = n \times m$ $n >> m$ is obtained by solving

$$A^TAx = A^Tb$$

Evaluating

$$\begin{bmatrix} G^T & \sqrt{\beta}\, W^T \end{bmatrix} \begin{bmatrix} G \\ \sqrt{\beta}\, W \end{bmatrix} m = \begin{bmatrix} G^T & \sqrt{\beta}\, W^T \end{bmatrix} \begin{bmatrix} d \\ \sqrt{\beta}\, W m_0 \end{bmatrix}$$

$$\boxed{(G^TG + \beta W^TW)\, m = G^Td + \beta W^TW m_0} \qquad (3)$$

is the same as (1). So the general quadratic optimization problem is

$$\min \quad \phi = \| W_d (Gm - d^{obs}) \|^2 + \beta \left\{ \alpha_s \| W_s (m - m_0) \|^2 + \alpha_x \| W_x (m - m_0) \|^2 \right\}$$

can be solved by finding the LS solution to the system

$$\boxed{\begin{bmatrix} W_d G \\ \sqrt{\beta \alpha_s}\, W_s \\ \sqrt{\beta \alpha_x}\, W_x \end{bmatrix} m = \begin{bmatrix} W_d d^{obs} \\ \sqrt{\beta \alpha_s}\, W_s m_0 \\ \sqrt{\beta \alpha_x}\, W_x m_0 \end{bmatrix}}$$

For large scale problem there are iterative solution techniques using conjugate gradients.