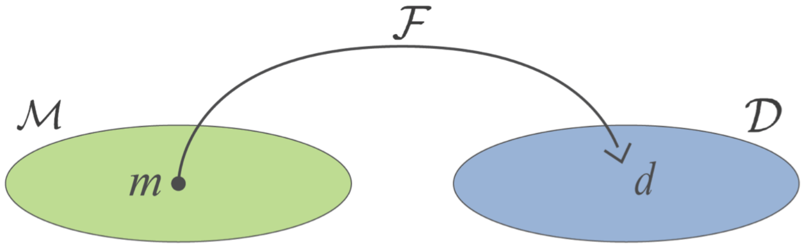
Singular Value Decomposition (SVD)

Outline

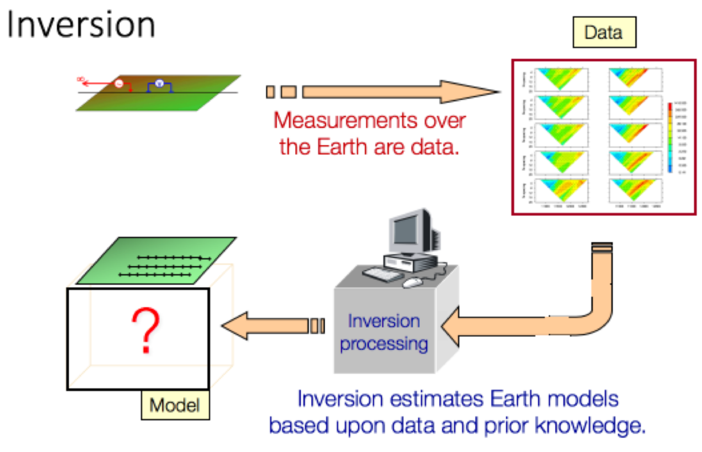
- Linear problem
- SVD (Singular Value Decomposition)
- Generalized inverse
- Controlling ill-conditioning
- Truncated SVD
- Understanding information from data alone

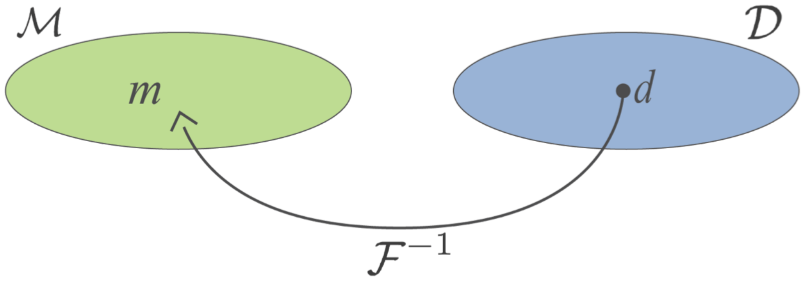
Our statement of the inverse problem

- Given observations: $d_j^{obs}, \quad j = 1, \dots, N$
  - Uncertainties: ϵ_j
  - Ability for forward modelling: $\mathcal{F}[m] = d$



- Find the earth model that gave rise to the data.



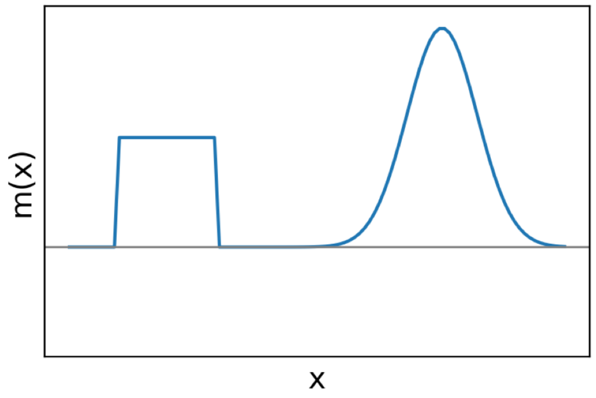


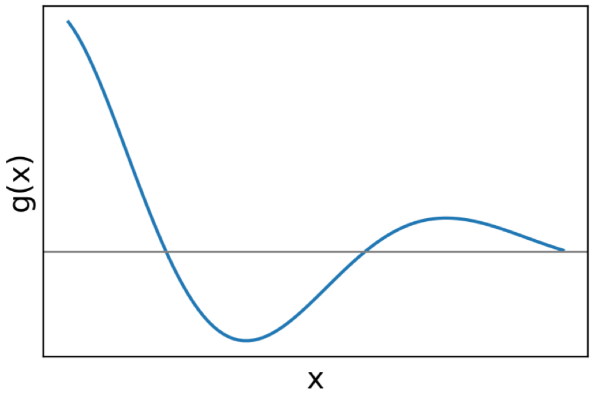
Linear problem

$$d_j = \int_v g_j(x) m(x) dx$$

$$g_j(x) = e^{jpx} cos(2\pi jqx)$$

- d_j : j-th datum
- g_j : kernel function for j-th datum
- m : model





Each datum in an inner product of the model with the kernel function. Evaluate one case

$$d_j = \mathbf{g} \cdot \mathbf{m} = 4.89$$

Basic challenges for the inverse problem:

Non-uniqueness

- Data are

$$d_i = \sum_{j=1}^{M} G_{ij} m_j$$

$$\mathbf{d} = \mathbf{G}\mathbf{m} \qquad \mathbf{G} : (N \times M)$$

- M>N (underdetermined problem) → infinitely many solutions
- Ill-conditioned

  - Small changes in the data yields large changes in the model

How do we understand and deal with these elements?

Solving $\mathbf{Gm} = \mathbf{d}$

Given $\mathbf{Gm} = \mathbf{d}$ $\quad \mathbf{G} \in \mathbb{R}^{N \times M} \quad \mathbf{m} \in \mathbb{R}^{M} \quad \mathbf{d} \in \mathbb{R}^{N}$

Want to write $\mathbf{m} = \mathbf{G}^{-1}\mathbf{d}$

$\mathbf{G}$: not sparse
$\mathbf{G}^{-1}$: does not exist

$\mathbf{m} = \mathbf{G}^{-1}\mathbf{d}$

Example m_1 + 2m_2 = 2
Write as Gm=d G=[1,2] d=2 m=(m1, m2)^T

Think m=G^-1 d

Use whatever language you use (e.g. matlab and python). What do you get

Singular Value Decomposition

$$\mathbf{G} = \mathbf{U}_p \boldsymbol{\Lambda} \mathbf{V}_p^T \qquad d_j = \int_v g_j(x) m(x) dx \qquad \mathbf{G} = \begin{pmatrix} — & \mathbf{g}_1 & — \\ & \vdots & \\ — & \mathbf{g}_N & — \end{pmatrix}$$

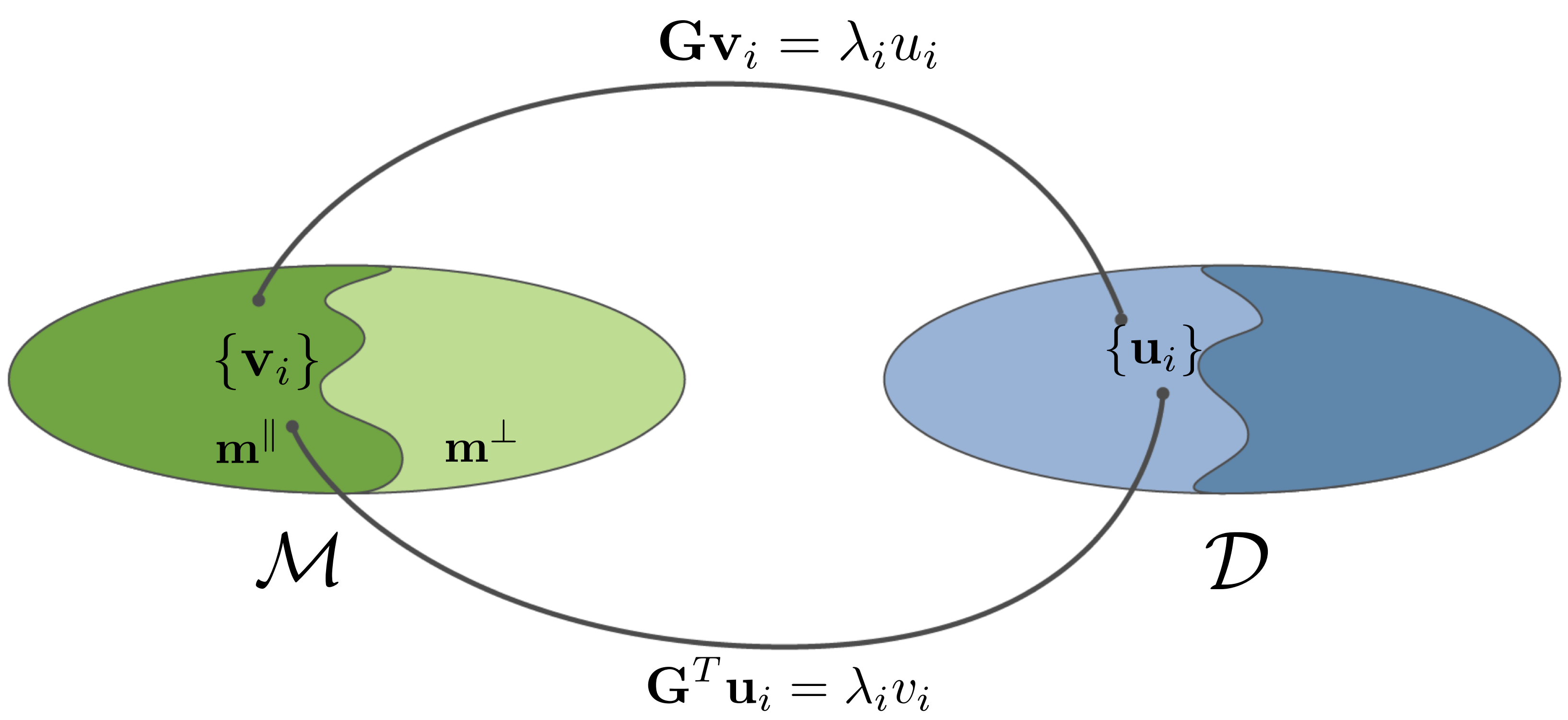
$$(M \times N)$$

$$\mathbf{U}_p = \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_p \\ | & & | \end{pmatrix} \quad \boldsymbol{\Lambda}_p = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix} \quad \mathbf{V}_p = \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \dots & \mathbf{v}_p \\ | & & | \end{pmatrix}$$

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

$$\begin{cases} \lambda_i\text{: singular values } (\lambda_1 \geq \lambda_2 \geq \dots \lambda_p > 0) \\ \mathbf{v}_i\text{: singular vectors (model space; } \mathbf{V}_p^T \mathbf{V}_p = \mathbf{I}_p) \\ \mathbf{u}_i\text{: singular vectors (data space; } \mathbf{U}_p^T \mathbf{U}_p = \mathbf{I}_p) \end{cases}$$

Singular Value Decomposition



$\mathbf{m}^{\parallel}$: activated portion of model space
$\mathbf{m}^{\perp}$:annihilator space

Solving with SVD

$$\mathbf{Gm} = \mathbf{d}$$

with $\mathbf{G} = \mathbf{U\Lambda V}^T$

$$\mathbf{U\Lambda V}^T\mathbf{m} = \mathbf{d}$$

$$\mathbf{\Lambda V}^T\mathbf{m} = \mathbf{U}^T\mathbf{d}$$

$$\mathbf{V}^T\mathbf{m} = \mathbf{\Lambda}^{-1}\mathbf{U}^T\mathbf{d}$$

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

Define $\mathbf{G}^\dagger = \mathbf{V\Lambda}^{-1}\mathbf{U}^T$

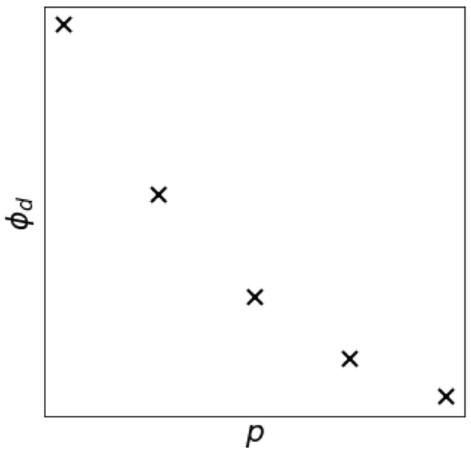
$$\mathbf{m}_c = \mathbf{G}^\dagger\mathbf{d}$$

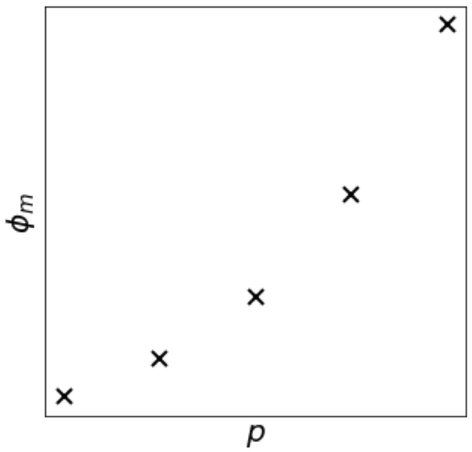
$$\mathbf{m}^\parallel = \mathbf{m}_c = \mathbf{VV}^T\mathbf{m} = \mathbf{V\Lambda}^{-1}\mathbf{U}^T\mathbf{d}$$

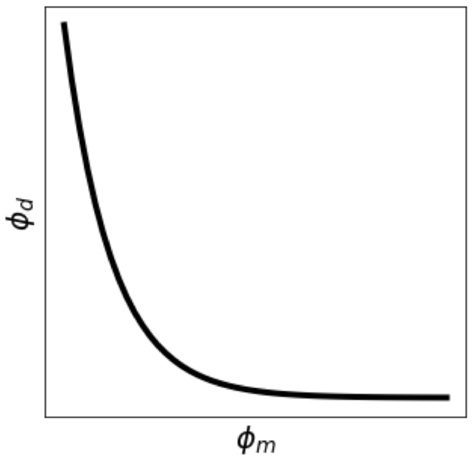
$$\boxed{\mathbf{m}_c = \mathbf{V\Lambda}^{-1}\mathbf{U}^T\mathbf{d} = \sum_{i=1}^{N}\left(\frac{\mathbf{u}_i^T\mathbf{d}}{\lambda_i}\right)\mathbf{v}_i}$$

Effects: Successively add vectors

Each vector $\mathbf{v}_i$ scaled by $\frac{\mathbf{u}_i^T\mathbf{d}}{\lambda_i} \to \infty$ if $\lambda_i \to 0$







$$\phi_d = \|\mathbf{Gm} - \mathbf{d}\|^2$$

$$\phi_m = \|\mathbf{m}\|^2$$

$$\mathbf{m}_c = \sum_{i=1}^{p}\left(\frac{\mathbf{u}_i^T\mathbf{d}}{\lambda_i}\right)\mathbf{v}_i$$

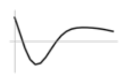
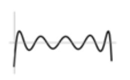
Ill-conditionedness

$$\mathbf{m}_c = \sum_{i=1}^{p} \left(\frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i} \right) \mathbf{v}_i$$

Small singular value

High oscillatory

Large amplitude of noise

λ_i	$\mathbf{v}_i$	$\mathbf{u}_i$	$\mathbf{u}_i^T \mathbf{d}^{true}$	$\frac{\mathbf{u}_i^T \mathbf{d}^{true}}{\lambda_i}$	$\mathbf{u}_i^T \mathbf{n}$	$\frac{\mathbf{u}_i^T \mathbf{n}}{\lambda_i}$
$\lambda_1 = 0.13$			-0.29	-2.13	-0.16	-1.16
$\lambda_2 = 0.11$			0.41	3.8	0.02	0.22
$\lambda_3 = 0.09$			0.16	1.9	0.08	0.96
$\lambda_{11} = 10^{-6}$			3.4×10^{-7}	0.32	-0.02	-1.9×10^4
$\lambda_{19} = 2.3 \times 10^{-17}$			-6.2×10^{-17}	-2.7	-0.0001	4.6×10^{13}

Truncated SVD

- If data are inaccurate, noise is also amplified by $1/\lambda_i$

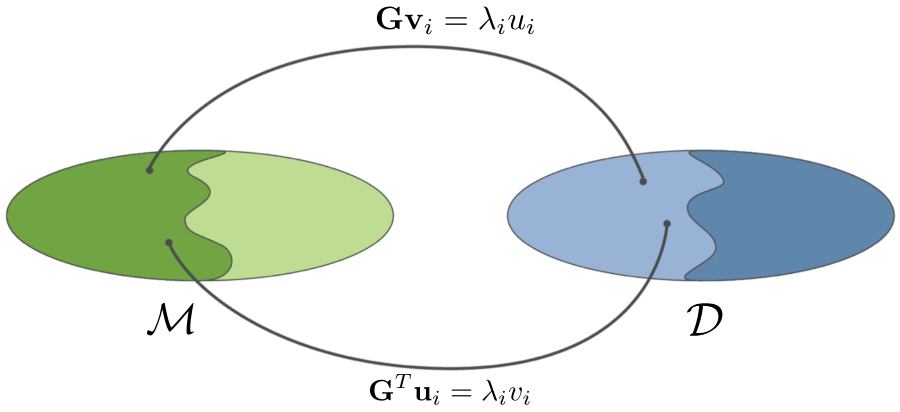
$$\mathbf{m}_c = \sum_{i=1}^{q} \left(\frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i}\right)\mathbf{v}_i + \underbrace{\sum_{i=q+1}^{N} \left(\frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i}\right)\mathbf{v}_i}_{\text{Cause more harm than good}}$$

- So

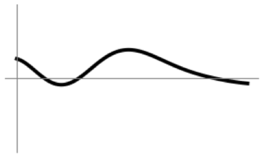
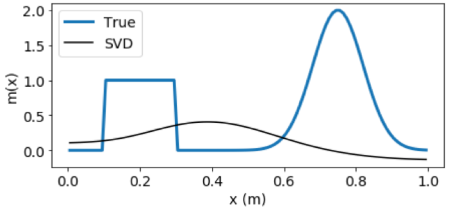
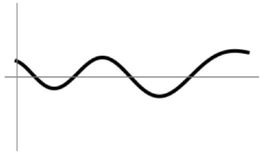
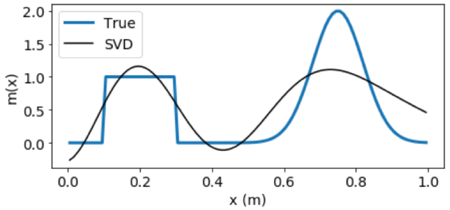
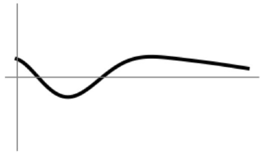
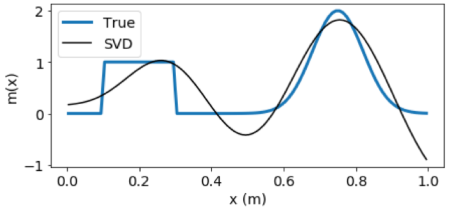
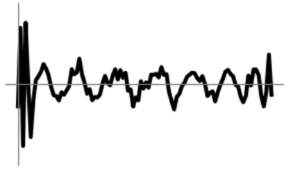
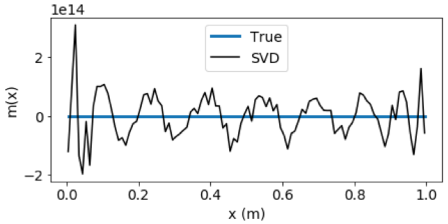
$$\mathbf{m}_c = \sum_{i=1}^{q} \left(\frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i}\right)\mathbf{v}_i$$

- Solution lies in a small sub-space

- Treats non-uniqueness and ill-conditioning

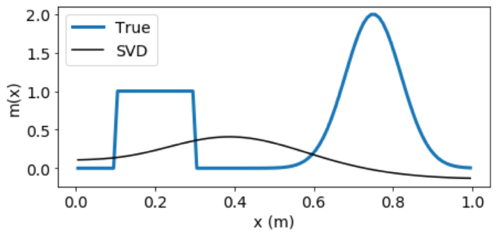


Truncated SVD

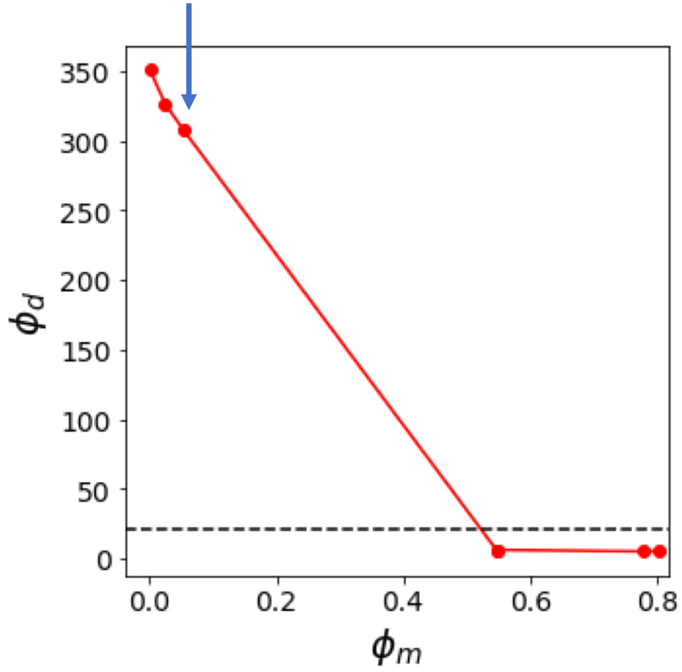
	$\mathbf{v}_i$	λ_i	$\mathbf{m}$		
$\mathbf{v}_3$		$\lambda_3 = 0.09$		$\mathbf{m}_c = \sum_{i=1}^{q} \left(\frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i}\right)\mathbf{v}_i$	$q = 3$
$\mathbf{v}_5$		$\lambda_5 = 0.04$		$\mathbf{m}_c = \sum_{i=1}^{q} \left(\frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i}\right)\mathbf{v}_i$	$q = 5$
$\mathbf{v}_6$		$\lambda_6 = 0.02$		$\mathbf{m}_c = \sum_{i=1}^{q} \left(\frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i}\right)\mathbf{v}_i$	$q = 6$
$\mathbf{v}_{19}$		$\lambda_{19} = 2.3 \times 10^{-17}$		$\mathbf{m}_c = \sum_{i=1}^{q} \left(\frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i}\right)\mathbf{v}_i$	$q = 18$

Plot Tikhonov curve

$q = 3$

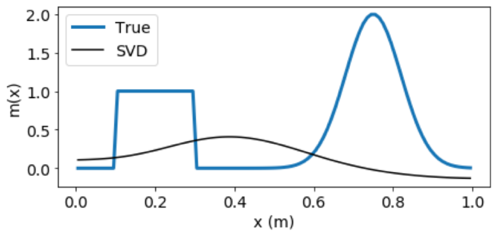


$q = 3$

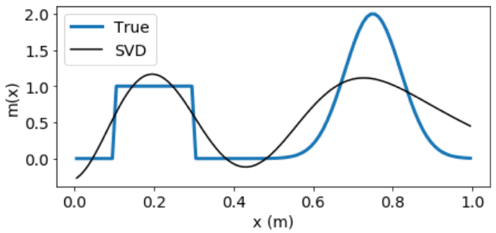


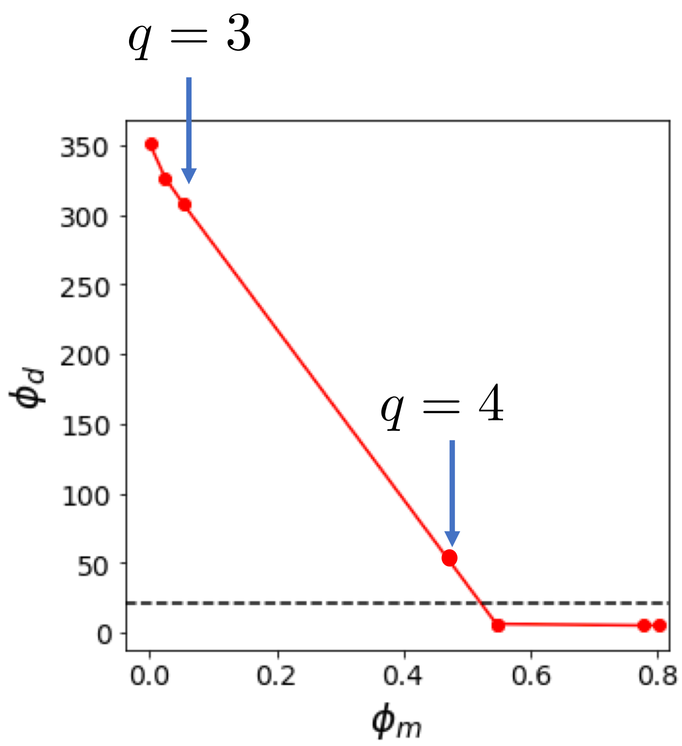
Plot Tikhonov curve

$q = 3$



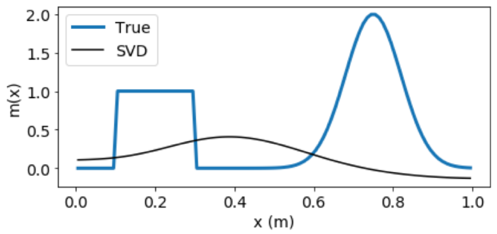
$q = 4$



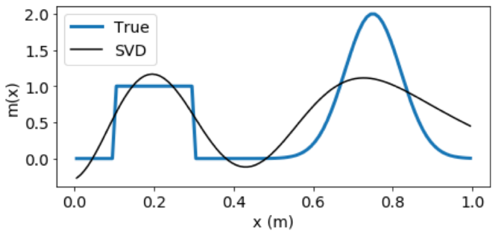


Plot Tikhonov curve

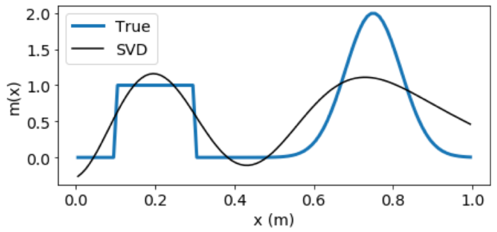
$q = 3$



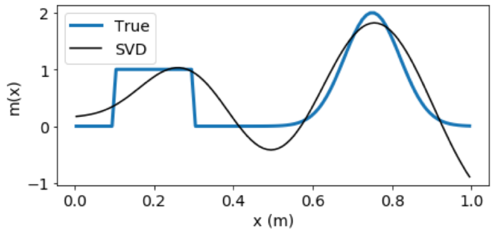
$q = 4$

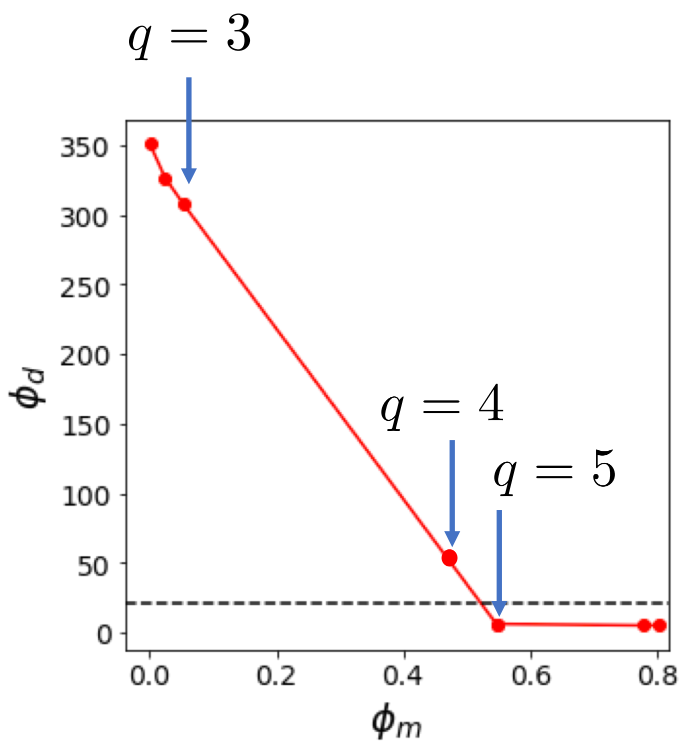


$q = 5$



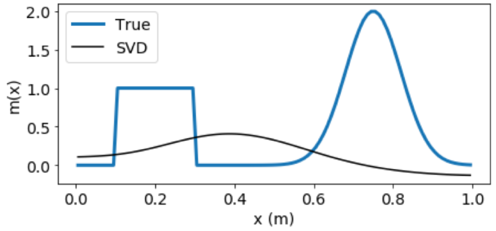
$q = 6$



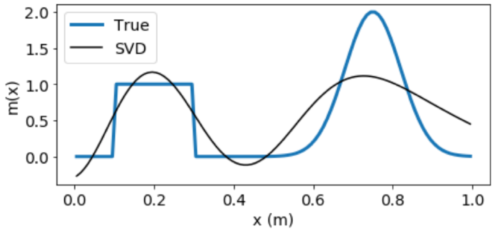


Plot Tikhonov curve

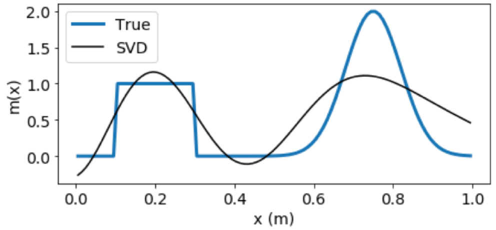
$q = 3$



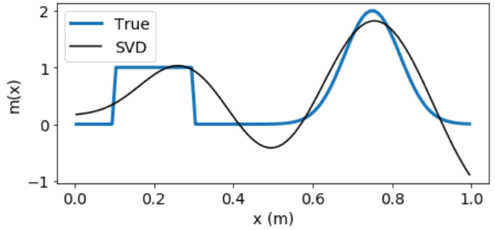
$q = 4$

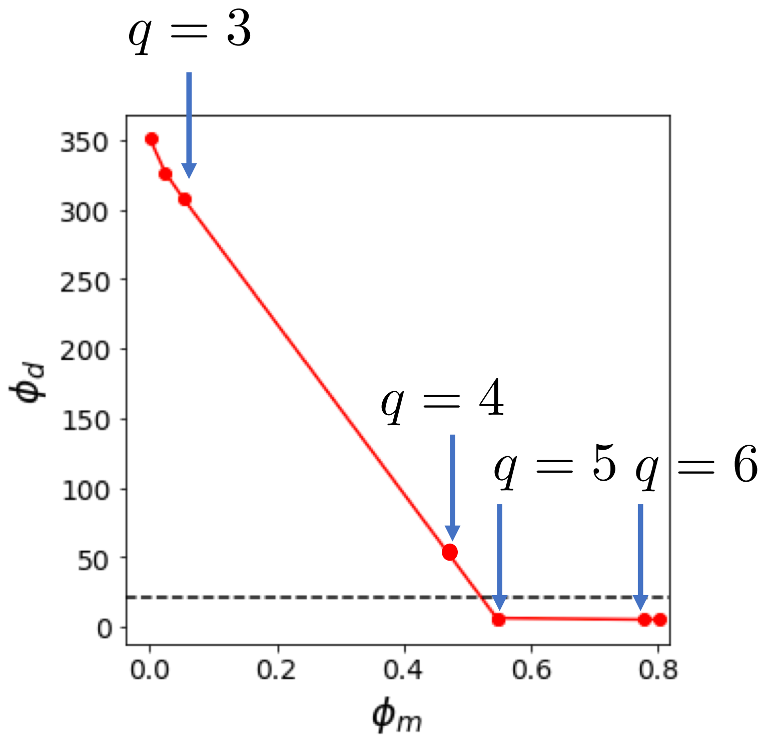


$q = 5$



$q = 6$





Consider Tikhonov solution

Data misfit: $\phi_d = \|\mathbf{Gm} - \mathbf{d}\|^2$

Regularization: $\phi_m = \|\mathbf{m}\|^2$

minimize $\phi = \phi_d + \beta\phi_m$

$$\phi = \|\mathbf{Gm} - \mathbf{d}\|^2 + \beta\|\mathbf{m}^c\|^2$$

minimum $\nabla \mathbf{g} = \nabla_m \phi = 0$

differentiate w.r.t vector $\dfrac{\partial \phi}{\partial \mathbf{m}} = 2\mathbf{G}^T\big(\mathbf{Gm} - \mathbf{d}\big) + 2\beta\mathbf{m} = 0$

So $\big(\mathbf{G}^T\mathbf{G} + \beta\mathbf{I}\big)\mathbf{m} = \mathbf{G}^T\mathbf{d}$

$\mathbf{G}^T\mathbf{G}$: $M \times M$ (rank deficient)

$\text{rank}(\mathbf{G}) \leq\ N,\ M > N$

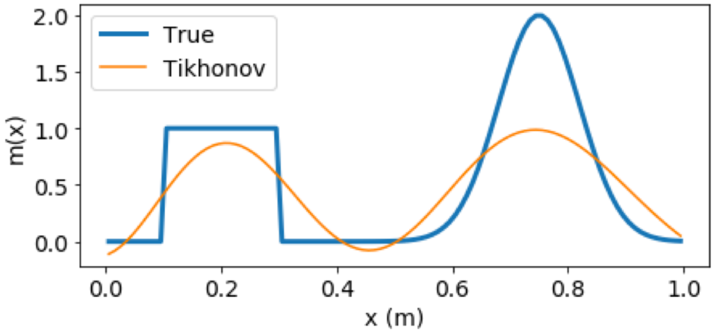
but $\beta > 0$, so adding $\beta\mathbf{I}$ makes all eigen values $\geq\ \beta$

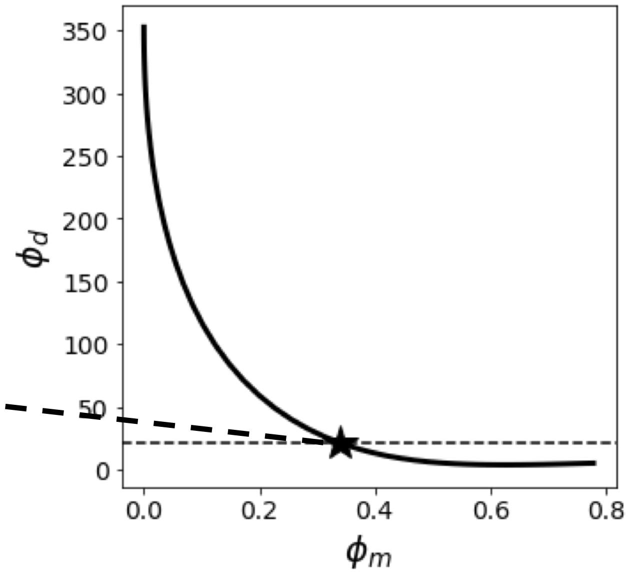
Solve $\mathbf{Am} = \mathbf{b}$ with letting $\mathbf{A} = \mathbf{G}^T\mathbf{G} + \beta\mathbf{I}$

$\mathbf{b} = \mathbf{G}^T\mathbf{d}$

Tikhonov curve

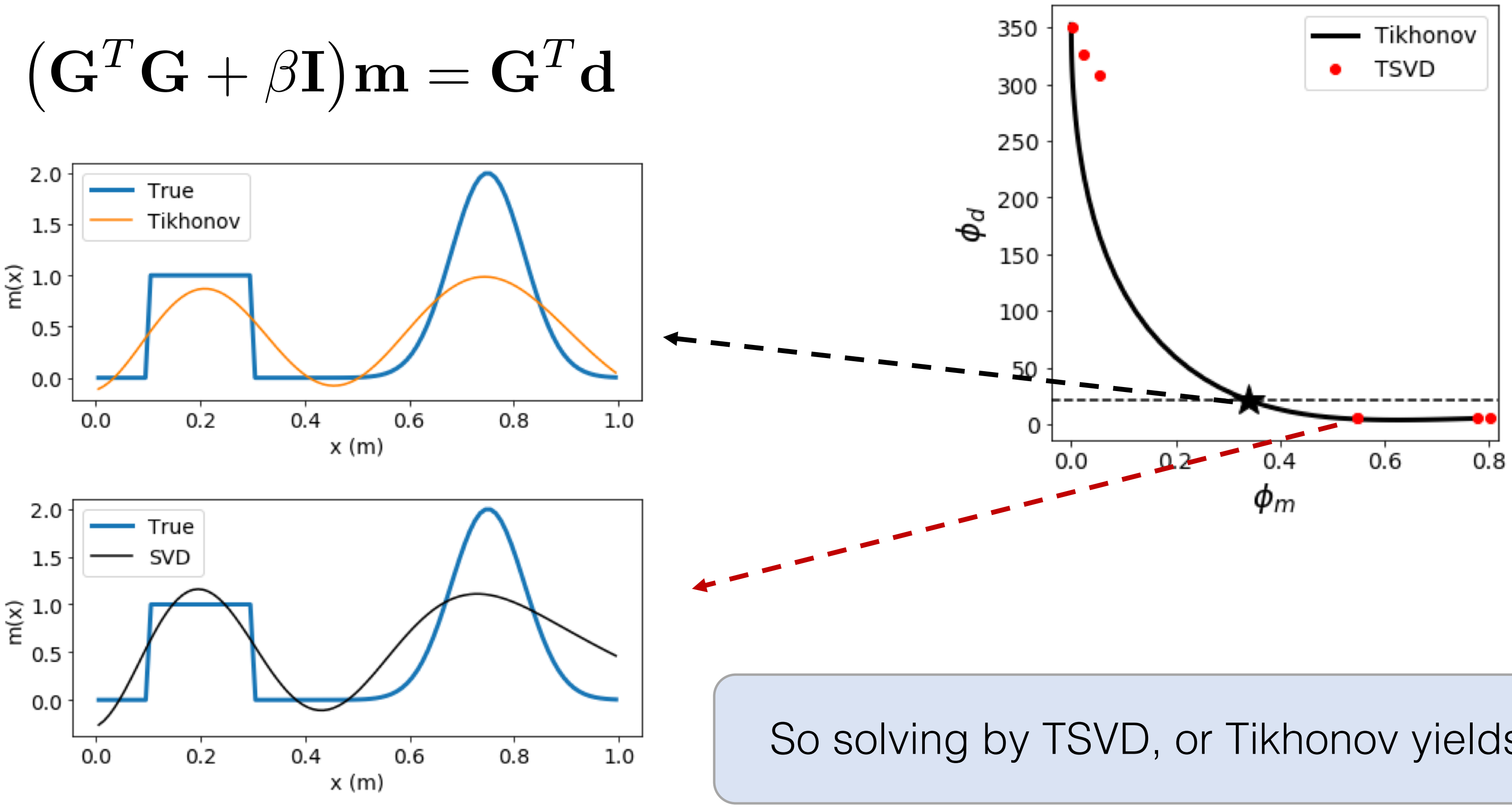
$$\left(\mathbf{G}^T\mathbf{G} + \beta\mathbf{I}\right)\mathbf{m} = \mathbf{G}^T\mathbf{d}$$





Tikhonov solution vs TSVD

$$(\mathbf{G}^T\mathbf{G} + \beta\mathbf{I})\mathbf{m} = \mathbf{G}^T\mathbf{d}$$



So solving by TSVD, or Tikhonov yields almost identical

Tikhonov, TSVD, and weighted TSVD

$$(\mathbf{G}^T\mathbf{G} + \beta\mathbf{I})\mathbf{m} = \mathbf{G}^T\mathbf{d}$$

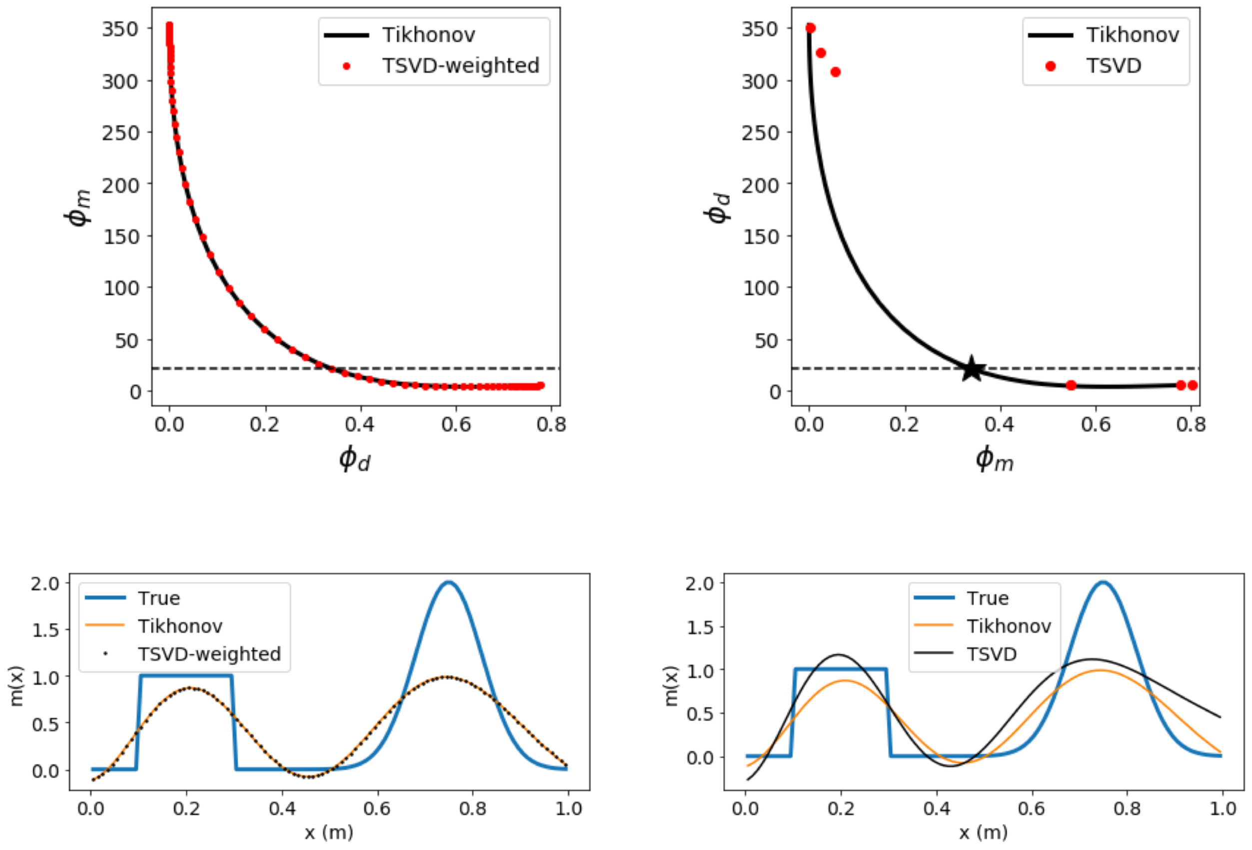
- In fact, exact correspondence is obtained by modifying to include weights

$$\mathbf{m}_c = \mathbf{V}\mathbf{T}\mathbf{\Lambda}^{-1}\mathbf{U}\mathbf{d}$$

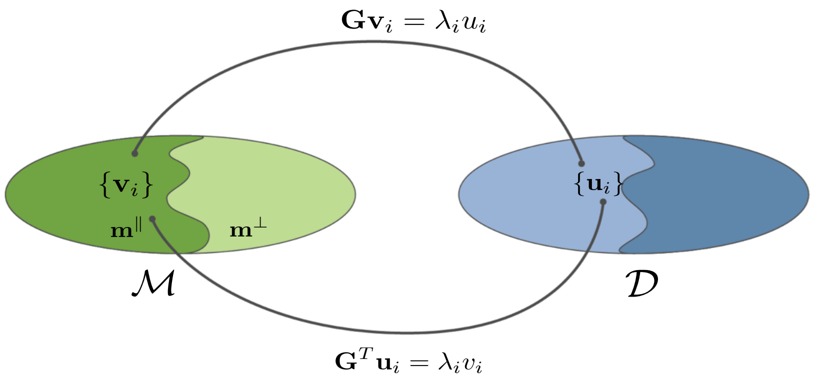
$$\mathbf{m}_c = \sum_{i=1}^{p} t_i \left(\frac{\mathbf{u}_i^T\mathbf{d}}{\lambda_i}\right)\mathbf{v}_i$$

$$\mathbf{T} = \mathbf{diag}(t_1, \ldots, t_p) \quad 0 \le t_i \le 1$$

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



Summary



$$\mathbf{G}\mathbf{m} = \mathbf{d}$$

$$\mathbf{G} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$$

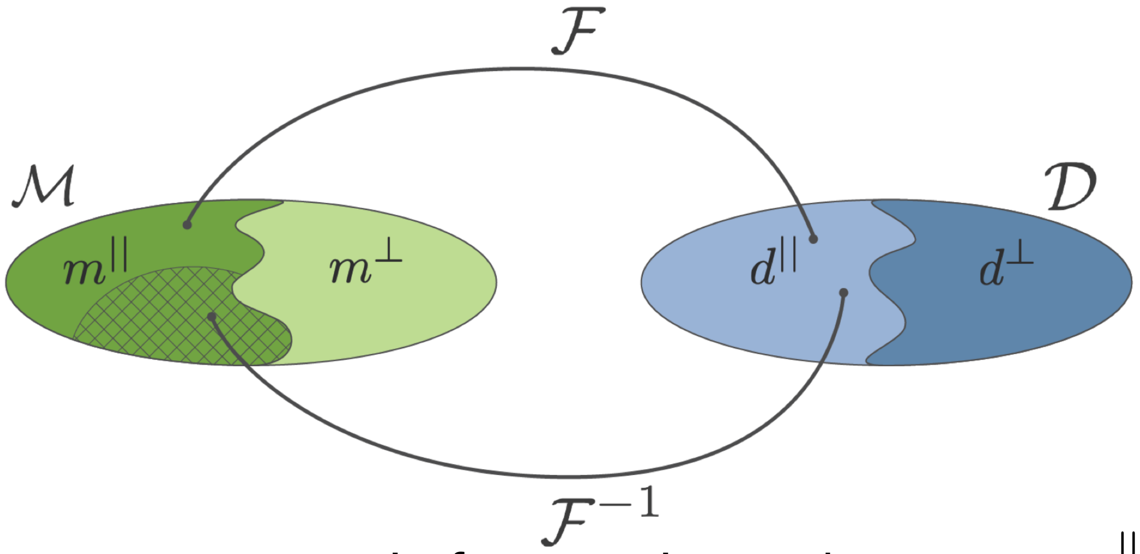
$$\mathbf{m}_c = \sum_{i=1}^{p} t_i \Big(\frac{\mathbf{u}_i^T\mathbf{d}}{\lambda_i}\Big)\mathbf{v}_i$$

- Small singular values: - Associated with high frequency vectors in model space
  - Amplify the noise
- Handling instabilities
  - (i) By truncating (keep only p)
  - (ii) Adjust their influence $t_i = \dfrac{\lambda_i}{\lambda_i^2 + \beta}$

this exactly the same as solving the Tikhonov problem

$$\begin{aligned}\text{minimize } \phi &= \phi_d + \beta\phi_m \\ &= \|\mathbf{G}\mathbf{m} - \mathbf{d}\|^2 + \beta\|\mathbf{m}\|^2\end{aligned}$$

What does the solution tell us



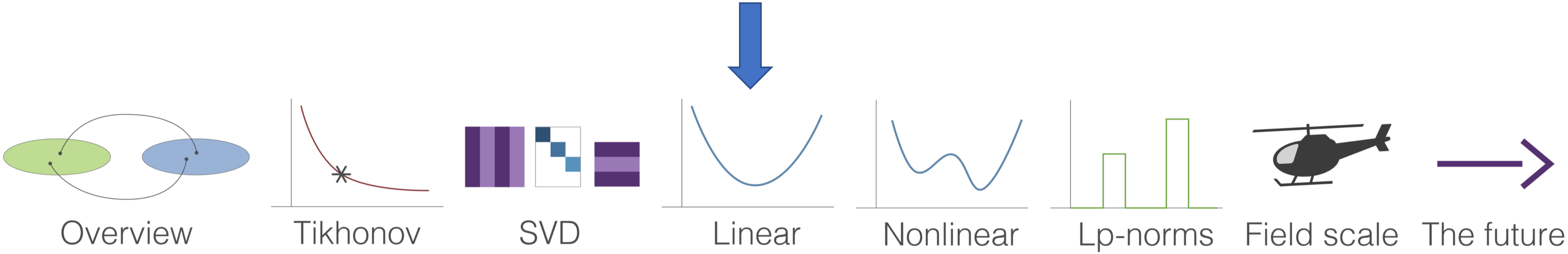
- Full data set can recover information about $\mathbf{m}^{||}$
- Regularized solution is in a reduced region of model space

Geophysical model lies outside of this region
To explore that we need to incorporate more information

Outline

- Linear problem
- SVD (Singular Value Decomposition)
- Generalized inverse
- Controlling ill-conditioning
- Truncated SVD
- Understanding information from data alone

Next up …


Overview
Tikhonov
SVD
Linear
Nonlinear
Lp-norms
Field scale
The future