1. Bardsley 2.1.

(a) Recall the minimization problem for Tikhonov regularization,

$$\boldsymbol{x}_\alpha = \arg\min_{\boldsymbol{x}} \left\{ \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 + \frac{\alpha}{2}\|\boldsymbol{x}\|^2 \right\}. \tag{2.5}$$

Define $\ell(x)$ to be the penalized least squares function and show that $\nabla\ell(\boldsymbol{x}) = \boldsymbol{0}$ yields

$$(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b} \tag{2.6}$$

and hence that solution of (2.5) and (2.6) coincide. Argue, moreover, that the solution is unique.

Solution:

Expanding $\ell(x)$ and denoting the inner product between two vectors as $\boldsymbol{x}\cdot\boldsymbol{y}$, we have

$$\begin{aligned} 2\ell(\boldsymbol{x}) &= (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \alpha\boldsymbol{x}^T\boldsymbol{x} \\ &= \boldsymbol{x}^T(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{x}^T(\boldsymbol{A}^T\boldsymbol{b}) + \boldsymbol{b}^T\boldsymbol{b} \\ &= \boldsymbol{x}^T(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{x} - 2\boldsymbol{x}\cdot(\boldsymbol{A}^T\boldsymbol{b}) + \boldsymbol{b}^T\boldsymbol{b} \quad \text{by symmetry of real inner products.} \end{aligned}$$

We evaluate the gradient $2\ell(\boldsymbol{x})$ term by term. The gradient of the constant $\boldsymbol{b}^t\boldsymbol{b}$ is $\boldsymbol{0}$, and the gradient of the vector function $\boldsymbol{x} \to \boldsymbol{x}\cdot(\boldsymbol{A}^T\boldsymbol{b})$ is precisely $\boldsymbol{A}^T\boldsymbol{b}$. Note that the quadratic form $f(\boldsymbol{x}) = \boldsymbol{x}^T(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{x}$ has

$$\begin{aligned} f(\boldsymbol{x} + \boldsymbol{h}) - f(\boldsymbol{x}) &= (\boldsymbol{x} + \boldsymbol{h})^T(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})(\boldsymbol{x} + \boldsymbol{h}) - \boldsymbol{x}^T(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{x} \\ &= \boldsymbol{h}^T(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{x} + \boldsymbol{x}^T(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{h} + \boldsymbol{h}^T(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{h} \\ &= 2\boldsymbol{h}\cdot\Big[(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{x} + (\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{h}\Big] \quad \text{by symmetry.} \end{aligned}$$

Note that if $\nabla f(\boldsymbol{x}) = 2(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{x}$, then

$$\lim_{\|\boldsymbol{h}\|\to 0} \frac{\|f(\boldsymbol{x} + \boldsymbol{h}) - f(\boldsymbol{x}) - \boldsymbol{h}\cdot\nabla f\|}{\|\boldsymbol{h}\|} = \lim_{\|\boldsymbol{h}\|\to 0} \frac{\|\boldsymbol{h}^T(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{h}\|}{\|\boldsymbol{h}\|} = 0$$

by Cauch-Schwartz applied twice. Hence, $\nabla\ell(\boldsymbol{x}) = 0$ if and only if

$$(\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{x} - \boldsymbol{A}^T\boldsymbol{b} + \boldsymbol{0} = \boldsymbol{0} \iff (\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I})\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}.$$

Recall that the singular value decomposition of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ is such that the diagonal elements of $\boldsymbol{\Sigma}$ are each non-negative real numbers. Moreover, if $\alpha > 0$, then $\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I} = \boldsymbol{V}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} + \alpha\boldsymbol{I})\boldsymbol{V}^T$ is an eignvalue decomposition with each eigenvalue strictly greater than 0 if and only if $\boldsymbol{A}^T\boldsymbol{A} + \alpha\boldsymbol{I}$ has full rank if and only if the solution to the matrix equation above is unique.

□

(b) Use the SVD of a matrix $\boldsymbol{A}$ to show that the solution of (2.6) (and hence (2.5)) can be written in filtered SVD form

$$\boldsymbol{x}_\alpha = \boldsymbol{V}\boldsymbol{\Phi}_\alpha\boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T\boldsymbol{b} \tag{2.3}$$

with Tikhonov filter

$$\boldsymbol{\Phi}_\alpha = \operatorname{diag}\left(\frac{\sigma_1^2}{\sigma_1^2+\alpha}, \ldots, \frac{\sigma_n^2}{\sigma_n^2+\alpha}\right). \tag{2.7}$$

Solution:

Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, then the Tikhonov solution is given by

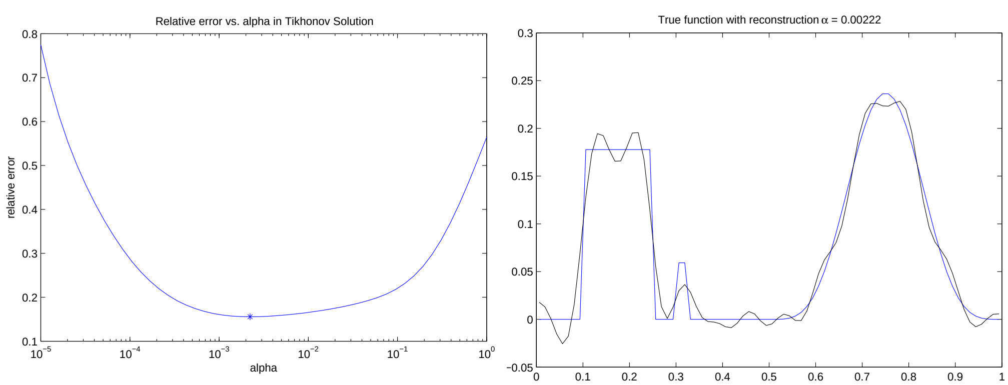
$$\begin{aligned}
\boldsymbol{v}_\alpha &= (\boldsymbol{A}^T\boldsymbol{A} - \alpha\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{b} \\
&= (\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T - \alpha\boldsymbol{V}\boldsymbol{V}^T)^{-1}\boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{U}^T\boldsymbol{b} \\
&= \boldsymbol{V}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} - \alpha\boldsymbol{I})^{-1}\cancel{\boldsymbol{V}^T\boldsymbol{V}}\boldsymbol{\Sigma}^T\boldsymbol{I}\boldsymbol{U}^T\boldsymbol{b} \\
&= \boldsymbol{V}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma} - \alpha\boldsymbol{I})^{-1}\boldsymbol{\Sigma}^T(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\dagger)\boldsymbol{U}^T\boldsymbol{b} \\
&= \boldsymbol{V}\boldsymbol{\Phi}_\alpha\boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T\boldsymbol{b}.
\end{aligned}$$

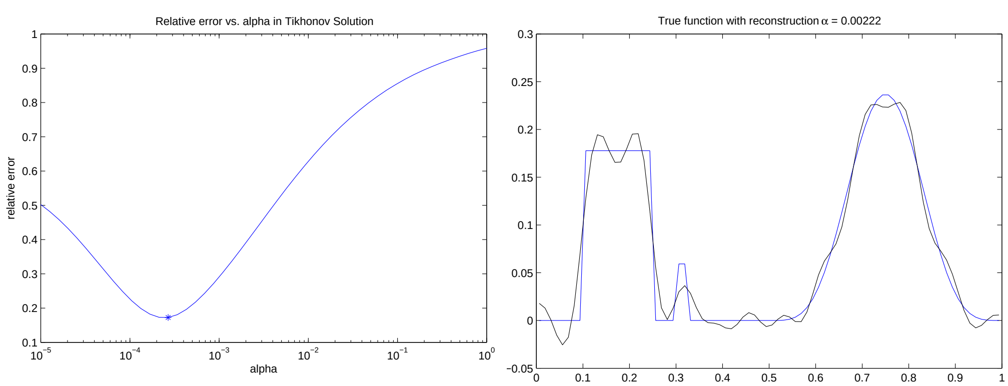
since $\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}$ and $\alpha\boldsymbol{I}$ are both $n \times n$ a diagonal matrices with diagonal elemenst σ_i^2 and α respectively, and $\boldsymbol{\Sigma}^\dagger$ is a right inverse for the matrix $\boldsymbol{\Sigma}$.

□

2. Bardsley 2.2.a. Create plots of the *relative error* function $f(\alpha) = \|\boldsymbol{x}_\alpha - \boldsymbol{x}_{\text{true}}\|/\|\boldsymbol{x}_{\text{true}}\|$ within both `DeblurTikhonov.m` and `PSFReconTikhonov.m` (use MATLAB's `logspace` function). Then estimate from the plots which α minimizes the relative error.

Solution:





$\square$

3. Bardsley 2.5.a & b

(a) Find the expressions for the MSE defined by

$$\text{MSE}(\nu) = \sigma^2 \sum_{i=1}^{n} (\phi_i^{(\nu)}/\sigma_i)^2 + \sum_{i=1}^{n} (1 - \phi_i^{(\nu)})^2 (\boldsymbol{v}_i^T \boldsymbol{x})^2. \tag{2.14}$$

for the TSVD, Tikhonov, and Landweber regularized solutions.

Solution:

In the case of TSVD, recall that the filter operator is the projection onto the first k singular vectors $v_1, \dots, v_k$. So $\phi_i^{(k)} = 1$ if $i \le k$ and 0 otherwise; hence,

$$\text{MSE}(k) = \sigma^2 \sum_{i=1}^{k} \sigma_i^{-2} + \sum_{i=k+1}^{n} (\boldsymbol{v}_i^T \boldsymbol{x})^2$$

For Tikhonov regularization, $\phi_i^{(\alpha)} = \sigma_i^2/(\sigma_i^2 + \alpha)$. Substituting into (2.14),

$$\text{MSE}(\alpha) = \sigma^2 \sum_{i=1}^{n} \left(\frac{\sigma_i}{\sigma_i^2 + \alpha} \right)^2 + \sum_{i=1}^{n} \left(\frac{\alpha}{\sigma_i^2 + \alpha} \right)^2 (\boldsymbol{v}_i^T \boldsymbol{x})^2.$$

For Landweber, recall $\phi_i^{(n,\tau)} = 1 - (1 - \tau\sigma_i^2)^n$ for $n = 1, 2, \dots$ and $0 < \tau < 1/\sigma_1$. Hence,

$$\begin{aligned} \text{MSE}(n,\tau) &= \sigma^2 \sum_{i=1}^{n} (\sigma_i^{-1}(1 - (1 - \tau\sigma_i^2)^n)^2 + \sum_{i=1}^{n} (1 - \tau\sigma_i^2)^{2n} (\boldsymbol{v}_i^T \boldsymbol{x})^2 \\ &= \sigma^2 \sum_{i=1}^{n} \left(\tau\sigma_i \sum_{k=0}^{n-1} (1 - \tau\sigma_i^2)^k \right)^2 + \sum_{i=1}^{n} (1 - \tau\sigma_i^2)^{2n} (\boldsymbol{v}_i^T \boldsymbol{x})^2. \end{aligned}$$

$\square$

(b) Create plots of MSE(α), using MATLAB's `logspace` function, within both `DeblurTikhonov.m` and `PSFreconTikhonov.m`. Then estimate from the plots which α minimizes the MSE.

This problem was combined with the following one.

4. Bardsley 2.6.a & d

(a) Modify `DeblurTikhonov.m` and `PSFreconTikhonov.m` so that (some subset of) the following curves are plotted together, (all on the same numerical grid created using `logspace`):

i the UPRE curve $U(\alpha)$ defined by

$$U(\alpha) = \sum_{i=1}^{b} \frac{\alpha^2 (\boldsymbol{u}_i^T \boldsymbol{b})^2}{(\sigma_i^2 + \alpha)^2} + 2\sigma^2 \sum_{i=1}^{n} \boldsymbol{\sigma_i^2} \sigma_i^2 + \alpha. \tag{2.21}$$

ii the GCV curve $G(\alpha)$ defined by

$$G(\alpha) = \left(\sum_{i=1}^{n} \frac{\alpha^2 (\boldsymbol{u}_i^T \boldsymbol{b})^2}{(\sigma_i^2 + \alpha)} \right) \Bigg/ \left(m - \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma_i^2 + \alpha} \right) \tag{2.25}$$

iii the DP curve $D(\alpha)^2$ defined by

$$D(\alpha) = \sum_{i=1}^{n} \frac{\alpha^2 (\boldsymbol{u}_i^T \boldsymbol{b})^2}{(\sigma_i^2 + \alpha)^2} - m\sigma^2 \tag{2.29}$$

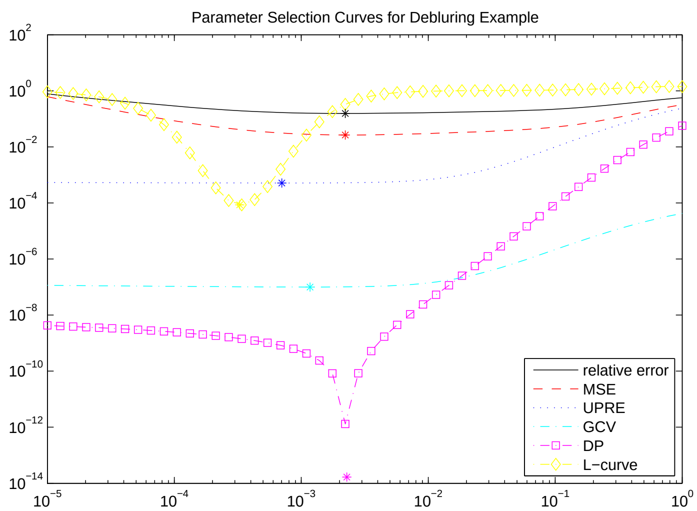
iv the L-curve curvature function $-C(\alpha)$ defined by

$$C(\alpha) = -\frac{r(\alpha)s(\alpha)[\alpha r(\alpha) + \alpha^2 s(\alpha)] + [r(\alpha)s(\alpha)]^2/s'(\alpha)]}{[r(\alpha)^2 + \alpha^2 s(\alpha)^2]^{3/2}} \tag{2.37}$$

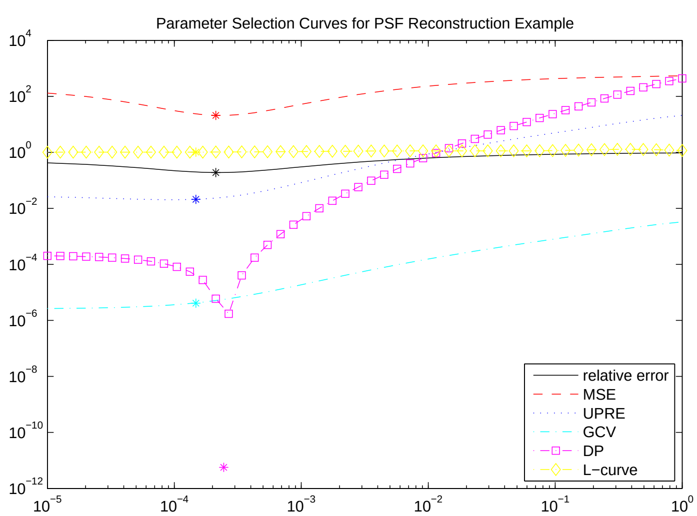
where $s(\alpha) = \|\boldsymbol{x}_\alpha\|^2$, and $r(\alpha) = \|\boldsymbol{A}\boldsymbol{x}_\alpha - \boldsymbol{b}\|^2$.

Solution:

Minimums for each curve were estimated using MATLABs `fminbnd` function. Note that each method selects a $\widehat{\alpha}$ that is bounded above by those given by the MSE and relative error.



relative error	$\widehat{\alpha} = 2.226 \times 10^{-3}$
MSE	$\widehat{\alpha} = 2.376 \times 10^{-3}$
UPRE	$\widehat{\alpha} = 7.024 \times 10^{-4}$
GCV	$\widehat{\alpha} = 1.173 \times 10^{-3}$
DP	$\widehat{\alpha} = 2.282 \times 10^{-3}$
L-curve	$\widehat{\alpha} = 3.271 \times 10^{-4}$



relative error	$\widehat{\alpha} = 2.518 \times 10^{-4}$
MSE	$\widehat{\alpha} = 2.185 \times 10^{-4}$
UPRE	$\widehat{\alpha} = 1.399 \times 10^{-4}$
GCV	$\widehat{\alpha} = 1.399 \times 10^{-4}$
DP	$\widehat{\alpha} = 1.828 \times 10^{-4}$
L-curve	$\widehat{\alpha} = 1.457 \times 10^{-4}$

□