Deep Learning and Inverse Problems
Summer 2024
Lecturer: Reinhard Heckel

Problem Set 3

Issued:	Tuesday, April	30, 2024,	1:00 pm,
Due:	Tuesday, May	7, 2024,	1:00 pm.

Problem 1 (Tikhonov-Regularized Least-Squares)**.** Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be an invertible matrix. Let $\mathbb{R}^n \ni \mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$ be a noisy measurement of a signal $\mathbf{x}^* \in \mathbb{R}^n$. Here, $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is a Gaussian noise term. In this problem, we discuss the regularized least-squares estimator

$$\hat{\mathbf{x}}_\lambda(\mathbf{y}) = \arg\min_{\mathbf{x}} ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2 + \lambda ||\mathbf{x}||_2^2, \tag{1}$$

of the true signal $\mathbf{x}^*$. Here, the factor $\lambda \geq 0$ is the regularization weight.
Hint: To solve this problem, it might be useful to review Section 3.3 from the lecture notes.

(a) Consider the singular-value decompositon $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ of the matrix $\mathbf{A}$. Here $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{n \times n}$ are orthonormal and $\mathbf{\Sigma}$ is a diagonal matrix with $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, ..., \sigma_n)$. Show that for all $\lambda \geq 0$ and for fixed $\mathbf{y}$, the vector

$$\hat{\mathbf{x}}_\lambda(\mathbf{y}) = \mathbf{V}\mathrm{diag}\left(\frac{\sigma_1}{\sigma_1^2 + \lambda}, ..., \frac{\sigma_n}{\sigma_n^2 + \lambda}\right)\mathbf{U}^T\mathbf{y}$$

is a solution of the regularized least-squares problem (1).

(b) Is the solution from (a) the unique minimizer? If yes, why? If no, state another minimizer.

(**Hint:** Strict convexity.)

(c) Show that the expected mean-squared error $\mathbb{E}_{\mathbf{e}}\left[||\hat{\mathbf{x}}_\lambda(\mathbf{y}) - \mathbf{x}^*||_2^2\right]$ (expectation is over the random noise $\mathbf{e}$) of the estimator $\hat{\mathbf{x}}_\lambda(\mathbf{y})$ satisfies

$$\mathbb{E}_{\mathbf{e}}\left[||\hat{\mathbf{x}}_\lambda(\mathbf{y}) - \mathbf{x}^*||_2^2\right] = \sum_{i=1}^n \left(1 - \frac{\sigma_i^2}{\sigma_i^2 + \lambda}\right)^2 (\mathbf{v}_i^T \mathbf{x}^*)^2 + \sigma^2 \sum_{i=1}^n \left(\frac{\sigma_i}{\sigma_i^2 + \lambda}\right)^2,$$

where $\mathbf{v}_i \in \mathbb{R}^n$ denotes the i-th column of the matrix $\mathbf{V}$.

Hint: First use the result from (a) to show that

$$\hat{\mathbf{x}}_\lambda(\mathbf{y}) = \mathbf{V}\mathrm{diag}\left(\frac{\sigma_1^2}{\sigma_1^2 + \lambda}, ..., \frac{\sigma_n^2}{\sigma_n^2 + \lambda}\right)\mathbf{V}^T\mathbf{x}^* + \mathbf{V}\mathrm{diag}\left(\frac{\sigma_1}{\sigma_1^2 + \lambda}, ..., \frac{\sigma_n}{\sigma_n^2 + \lambda}\right)\mathbf{U}^T\mathbf{e}.$$

(d) Assume you know a good regularization weight $\bar{\lambda}$ for the noise $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \bar{\sigma}^2)$. If we change the noise model to $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \tilde{\sigma}^2)$ with $\tilde{\sigma}^2 > \bar{\sigma}^2$, how would you adapt the regularization weight λ? Would you make it larger or smaller than $\bar{\lambda}$? Justify your answer!

Problem 1 (Tikhonov-Regularized Least-Squares)**.** Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be an invertible matrix. Let $\mathbb{R}^n \ni \mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$ be a noisy measurement of a signal $\mathbf{x}^* \in \mathbb{R}^n$. Here, $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ is a Gaussian noise term. In this problem, we discuss the regularized least-squares estimator

$$\hat{\mathbf{x}}_\lambda(\mathbf{y}) = \arg\min_{\mathbf{x}} ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2 + \lambda\,||\mathbf{x}||_2^2\,, \tag{1}$$

of the true signal $\mathbf{x}^*$. Here, the factor $\lambda \geq 0$ is the regularization weight.
Hint: To solve this problem, it might be useful to review Section 3.3 from the lecture notes.

(a) Consider the singular-value decompositon $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ of the matrix $\mathbf{A}$. Here $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{n\times n}$ are orthonormal and $\mathbf{\Sigma}$ is a diagonal matrix with $\mathbf{\Sigma} = \text{diag}(\sigma_1, ..., \sigma_n)$. Show that for all $\lambda \geq 0$ and for fixed $\mathbf{y}$, the vector

$$\hat{\mathbf{x}}_\lambda(\mathbf{y}) = \mathbf{V}\text{diag}\left(\frac{\sigma_1}{\sigma_1^2+\lambda}, ..., \frac{\sigma_n}{\sigma_n^2+\lambda}\right)\mathbf{U}^T\mathbf{y}$$

is a solution of the regularized least-squares problem (1).

$$y = Ax^* + e \quad , \quad e \sim \mathcal{N}(0, \sigma^2 I)$$

for LS:

$$\hat{x}_{LS} = A^{-1}y \implies \hat{x}_{LS} = V\Sigma^{-1}U^T y$$

Regularization:

$$\hat{x}_\lambda(y) = \underset{x}{\text{argmin}}\ \|Ax-y\|_2^2 + \lambda\|x\|_2^2$$

$$= (A^TA + \lambda I)^{-1}A^T y$$

$$= (A^TA)^{-1}A^T y + \frac{1}{\lambda}I^{-1}A^T y$$

$$= A^{-1}y + \frac{1}{\lambda}A^T y = (A^{-1} + \frac{1}{\lambda}A^T)y$$

$$\boxed{\begin{aligned} \Rightarrow A &= U\Sigma V^T \\ A^T &= V\Sigma U^T \\ A^{-1} &= V\Sigma^{-1}U^T \end{aligned}}$$

$$= (V\Sigma^{-1}U^T + \frac{1}{\lambda}V\Sigma U^T)y$$

$$= V\underbrace{(\Sigma^{-1} + \frac{1}{\lambda}\Sigma)}_{(*)}U^T y$$

$$(*) \Rightarrow \operatorname{diag}\left(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_n}\right) + \operatorname{diag}\left(\frac{\sigma_1}{\lambda}, \dots, \frac{\sigma_n}{\lambda}\right)$$

$$= \operatorname{diag}\left(\frac{1}{\sigma_1} + \frac{\sigma_1}{\lambda}, \dots, \frac{1}{\sigma_n} + \frac{\sigma_n}{\lambda}\right)$$

$$= \operatorname{diag}\left(\frac{\lambda + \sigma_1^2}{\sigma_1 \lambda}, \dots, \frac{\lambda + \sigma_n^2}{\sigma_n \lambda}\right)$$

$$\hat{x}_\lambda(y) = V \operatorname{diag}\underbrace{\left(\frac{\lambda + \sigma_1^2}{\sigma_1 \lambda}, \dots, \frac{\lambda + \sigma_n^2}{\sigma_n \lambda}\right)}_{\text{I found it slightly different.}} U^T y$$

$$\lambda(*)^{-1} = \text{Correct answer?}$$

(b) Is the solution from (a) the unique minimizer? If yes, why? If no, state another minimizer. (**Hint:** Strict convexity.)

If the optimization problem is convex, then it has one unique minimum.

Convex Optimization: convex cost
affine equality constraints
convex inequality — " —

1) It doesn't have constraints. ✓

2) Convex cost?
Check if $\|Ax - y\|_2^2 + \lambda \|x\|_2^2$ is convex.

Since it is differentiable, let's check the Hessian.

$f(x) = \|Ax-y\|_2^2 + \lambda\|x\|_2^2 = (Ax-y)^T(Ax-y) + \lambda x^T x$

$\frac{\partial f(x)}{\partial x} = 2(Ax-y)^T A + 2\lambda x^T$

$= 2x^T A^T A - 2y^T A + 2\lambda x^T$

$\frac{\partial^2 f(x)}{\partial x^2} = \underbrace{2(A^T A)^T}_{\succ 0} + \underbrace{2\lambda \mathbb{1}}_{\succ 0}$, where $\mathbb{1}$ is Identity matrix

$A^T A$ is always positive definite

$\lambda \mathbb{1}$ is already positive definite with $\lambda > 0$ and positive semi-definite for $\lambda \geq 0$, so the Hessian is positive definite or at least positive semi-definit, so the function is convex.

Which means that the optimization is a convex optimization which has one unique solution globally.

(c) Show that the expected mean-squared error $\mathbb{E}_{\mathbf{e}}\left[||\hat{\mathbf{x}}_\lambda(\mathbf{y}) - \mathbf{x}^*||_2^2\right]$ (expectation is over the random noise $\mathbf{e}$) of the estimator $\hat{\mathbf{x}}_\lambda(\mathbf{y})$ satisfies

$$\mathbb{E}_{\mathbf{e}}\left[||\hat{\mathbf{x}}_\lambda(\mathbf{y}) - \mathbf{x}^*||_2^2\right] = \sum_{i=1}^{n}\left(1 - \frac{\sigma_i^2}{\sigma_i^2+\lambda}\right)^2 (\mathbf{v}_i^T\mathbf{x}^*)^2 + \sigma^2\sum_{i=1}^{n}\left(\frac{\sigma_i}{\sigma_i^2+\lambda}\right)^2,$$

where $\mathbf{v}_i \in \mathbb{R}^n$ denotes the i-th column of the matrix $\mathbf{V}$.

Hint: First use the result from (a) to show that

$$\hat{\mathbf{x}}_\lambda(\mathbf{y}) = \mathbf{V}\mathrm{diag}\left(\frac{\sigma_1^2}{\sigma_1^2+\lambda}, ..., \frac{\sigma_n^2}{\sigma_n^2+\lambda}\right)\mathbf{V}^T\mathbf{x}^* + \mathbf{V}\mathrm{diag}\left(\frac{\sigma_1}{\sigma_1^2+\lambda}, ..., \frac{\sigma_n}{\sigma_n^2+\lambda}\right)\mathbf{U}^T\mathbf{e}.$$

Hint: $\hat{x}_\lambda(y) = V \operatorname{diag}\left(\frac{\sigma_1}{\sigma_1^2+\lambda}, \dots, \frac{\sigma_n}{\sigma_n^2+\lambda}\right) U^T y$

$$y = Ax^* + e$$
$$= U\Sigma V^T x^* + e$$

$$\hat{x}_\lambda(y) = V \operatorname{diag}\left(\frac{\sigma_1}{\sigma_1^2+\lambda}, \dots\right) \underbrace{U^T U}_{I} \Sigma V^T x^* + V \operatorname{diag}\left(\frac{\sigma_1}{\sigma_1^2+\lambda}, \dots\right) U^T e$$

$$= V \operatorname{diag}\left(\frac{\sigma_1^2}{\sigma_1^2+\lambda}, \dots\right) V^T x^* + V \operatorname{diag}\left(\frac{\sigma_1}{\sigma_1^2+\lambda}, \dots\right) U^T e$$

introduce $S = \operatorname{diag}\left(\frac{\sigma_1}{\sigma_1^2+\lambda}, \dots, \frac{\sigma_n}{\sigma_n^2+\lambda}\right)$

$$\hat{x}_\lambda(y) = VS\Sigma V^T x^* + VSU^T e$$

$$E_e\left[\|\hat{x}_\lambda(y) - x^*\|_2^2\right] = E_e\left[(\hat{x}_\lambda(y) - x^*)^T(\hat{x}_\lambda(y) - x^*)\right] =$$

$$= E_e\left[\hat{x}_\lambda(y)^T x_\lambda(y) - 2\hat{x}_\lambda(y)^T x^* + x^{*T} x^*\right]$$

$$= E_e\left[\hat{x}_\lambda(y)^T \hat{x}_\lambda(y)\right] + -2E\left[\hat{x}_\lambda(y)^T x^*\right] + \underbrace{E\left[x^{*T} x^*\right]}_{=0}$$

$$= \overbrace{E_e\left[(VS\Sigma V^T x^* + VSU^T e)^T (VS\Sigma V^T x^* + VSU^T e)\right]}^{I} +$$

$$\underbrace{-2E\left[(VS\Sigma V^T x^* + VSU^T e)^T x^*\right]}_{II}$$

I) Get rid of the zero terms

$$= (VS\Sigma V^T x^*)^T VS\Sigma V^T x^* + 2\mathbb{E}_e\Big[\underbrace{(VS\Sigma V^T x^*)^T VSU^T e}_{=0}\Big]$$

$$+\mathbb{E}_e\big[(VSU^T e)^T VSU^T e\big]$$

$$= x^{*T} V\Sigma S V^T V S \Sigma V^T x^* + \mathbb{E}_e\big[e^T U S V^T V S U^T e\big]$$

$$= x^{*T} V \Sigma^2 S^2 V^T x^* + \mathbb{E}_e\big[e^T U S^2 U^T e\big]$$

II) $= -2\, x^{*T} V\Sigma S V^T x^{*T} + \underbrace{\mathbb{E}_e[e^T \ldots]}_{=0}$

I)+II) $\Rightarrow$ $= x^{*T} V\big(\Sigma^2 S^2 - 2\Sigma S\big) V^T x^* + \mathbb{E}_e\big[e^T U S^2 U^T e\big]$

$$\Sigma S = \mathrm{diag}\left(\frac{\sigma_1^2}{\sigma_1^2+\lambda}, \ldots, \frac{\sigma_n^2}{\sigma_n^2+\lambda}\right)$$

$$(\Sigma S)^2 = \mathrm{diag}\left(\left(\frac{\sigma_1^2}{\sigma_1^2+\lambda}\right)^2, \ldots, \left(\frac{\sigma_n^2}{\sigma_n+\lambda}\right)^2\right)$$

$$(\Sigma S)^2 - \Sigma S = \mathrm{diag}\left(\frac{-\sigma_1^2\lambda}{(\sigma_1^2+\lambda)^2}, \ldots, \frac{-\sigma_n^2\lambda}{(\sigma_n^2+\lambda)^2}\right)$$

$$\frac{\sigma_1^4}{(\sigma_1^2+\lambda)^2} - \frac{\sigma_1^2}{\sigma_1^2+\lambda} =$$

$$= \frac{\sigma_1^4 - \sigma_1^2(\sigma_1^2+\lambda)}{(\sigma_1^2+\lambda)^2}$$

$$= \frac{-\sigma_1^2\lambda}{(\sigma_1^2+\lambda)^2}$$

I) + II) $\Rightarrow x^{*T} V \operatorname{diag}\left(\frac{-\sigma_1^2 \lambda}{(\sigma_1^2+\lambda)^2}, \dots, \frac{-\sigma_n \lambda}{(\sigma_n^2+\lambda)^2}\right) V^T x^* + \mathbb{E}\left[e^T S^2 e\right]$

U doesn't change the mean

$$= \underbrace{\sum_{i=1}^{n} \left(\frac{-\sigma_i^2 \lambda}{(\sigma_i^2+\lambda)^2}\right) (V^T x_*)^2}_{\text{not really}} + \underbrace{\sum_{i=1}^{n} \left(\frac{\sigma_i}{\sigma_i^2+\lambda}\right)^2 \cdot \sigma^2}_{\text{correct}}$$

Variance of e

$E[e^T e] = \sigma^2$

d) Assume you know a good regularization weight $\bar{\lambda}$ for the noise $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \bar{\sigma}^2)$. If we change the noise model to $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \tilde{\sigma}^2)$ with $\tilde{\sigma}^2 > \bar{\sigma}^2$, how would you adapt the regularization weight λ? Would you make it larger or smaller than $\bar{\lambda}$? Justify your answer!

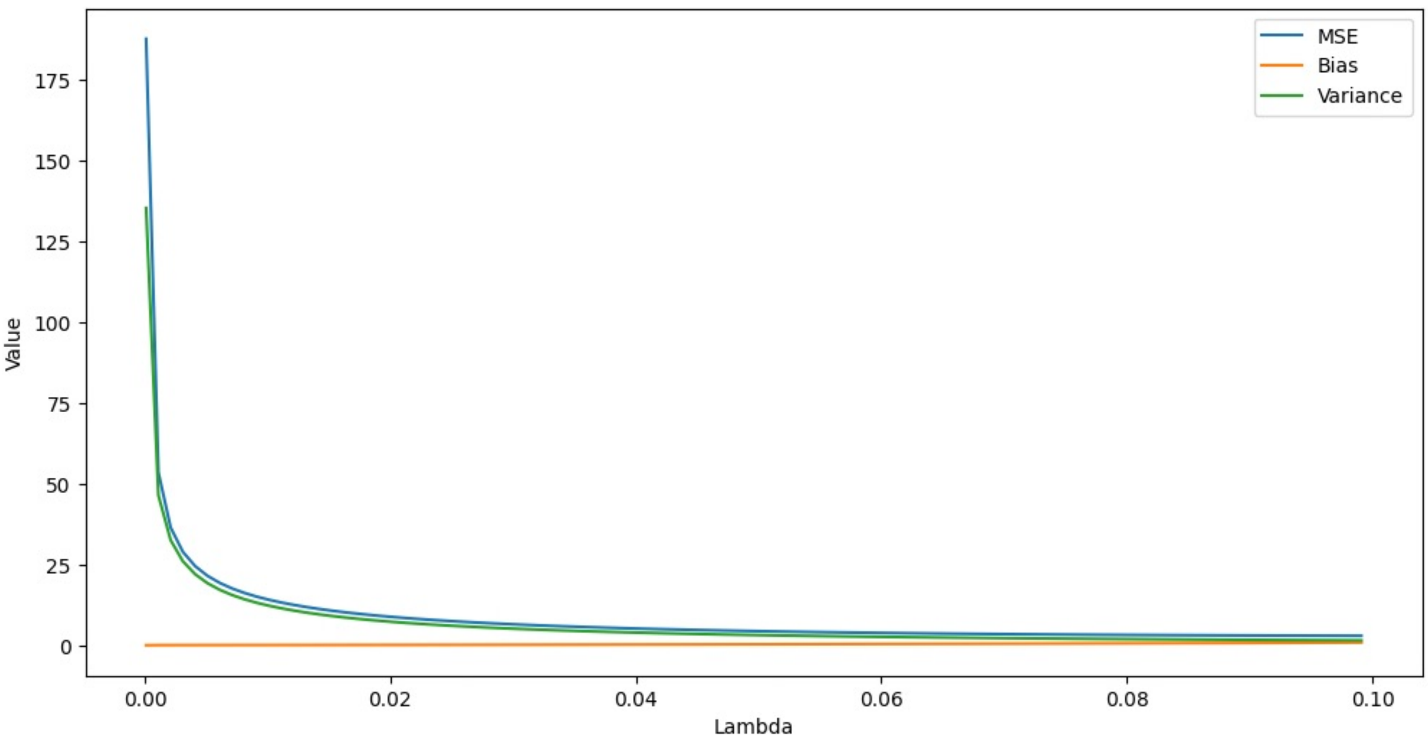
$$\mathbb{E}_{\mathbf{e}}\left[||\hat{\mathbf{x}}_\lambda(\mathbf{y}) - \mathbf{x}^*||_2^2\right] = \sum_{i=1}^{n} \left(1 - \frac{\sigma_i^2}{\sigma_i^2 + \lambda}\right)^2 (\mathbf{v}_i^T \mathbf{x}^*)^2 + \sigma^2 \sum_{i=1}^{n} \left(\frac{\sigma_i}{\sigma_i^2 + \lambda}\right)^2,$$

$\tilde{\sigma}^2 > \bar{\sigma}^2$

If the variance increases, then the expectation of the difference increases as well, which means that the estimation becomes bad. In order to reduce the effect of variance, the term $\left(\frac{\sigma_i}{\sigma_i^2+\lambda}\right)$ has to be smaller. Therefore we have to increase λ. $\tilde{\lambda} > \bar{\lambda}$

Problem 2 (Regularizing Deconvolution)**.** Here, we numerically solve the deblurring problem discussed in the lecture notes. For simplicity, we consider the 1D case. Let $\mathbf{x}^* \in \mathbb{R}^n$ be a 1D signal. Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be the matrix that implements convolution with the Gaussian kernel from Figure 3.1 of the notes. We want to use Tikhonov reulgarization to recover the signal $\mathbf{x}^*$ from a noisy measurement $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$, where $\mathbf{e}$ is Gaussian noise. In this problem, your task is to write code to reproduce the bias-variance tradeoff (Figure 3.1 of the lecture notes). To that end, implement the matrix $\mathbf{A}$ to carry out the Gaussian convolution and make use of the results derived in problem 1. **Hint:** Feel free to use generative AI models such as Copilot or ChatGPT to help you with the coding.

For $\sigma = 0,01$ //



Bias Values are relatively small [0,01 ; 0,89]