Math 164 - Bonus Assignment, Fall 2024

- **Due Date:** 11:59 pm on December 6th, 2024

This bonus assignment considers the reconstruction of an image from a set of parallel projections, acquired along different angles. Such a dataset is acquired in *computed tomography* (CT).

The acquisition process can be seen as a *linear transformation* of an image of size $l \times l$ pixels – in which we are interested, but *don't have direct access to* – to another image of size $n \times l$ called a *sinogram*. Please check out this video to get some intuition.

Since this measurement process is linear, we can write the ideal data acquisition as

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^{nl}$ is the vectorized sinogram image, $\mathbf{x} \in \mathbb{R}^{l^2}$ is the vectorized image we are interested in recovering, and $\mathbf{A} \in \mathbb{R}^{nl \times l^2}$ is the image formation operator. In practice there is noise in the measurements. So a more accurate *forward model* would be

$$\mathbf{y}_\eta = \mathbf{A}\mathbf{x} + \eta, \tag{2}$$

where $\eta \in \mathbb{R}^{nl}$ is the measurement noise.

The goal for this bonus assignment is for you to get some intuition into: (i) the difficulties of solving the *inverse problem* of finding $\mathbf{x}$ from $\mathbf{y}_\eta$, (ii) different ways of *modeling* a suitable optimisation problem, and (iii) what optimization algorithms to use to solve the problem. In the corresponding .ipynb file you'll find code for generating the matrix $\mathbf{A}$ and the ground truth $\mathbf{x}$ that we will try to reconstruct. Use $l = 128$ throughout this assignment.

1. **Difficulties of inverse problems.** Our goal here is to understand why it will be tricky to reconstruct an image from data. For that, we will consider the role of the number of measurement directions n and the noise η:

   (a) If $n = l$ and the the matrix $\mathbf{A}$ were invertible, we could consider $\hat{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{y}_\eta$ as an estimate for the reconstruction. However, in practice folks take (for various reasons) $n < l$ directions into account. Compute and plot the eigenvalues of $\mathbf{A}^\top\mathbf{A}$ for $n \in \{l/8, l/4, l/2, l\}$ and describe what you see.

   (b) Instead of using the inverse of $\mathbf{A}$, consider the *pseudo-inverse* given by $\mathbf{A}^\dagger := \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}$. Compute and plot the reconstruction $\mathbf{A}^\dagger\mathbf{y}_\eta \in \mathbb{R}^{l^2}$ as an $l \times l$ image for data generated with noise $\eta \sim \mathcal{N}(0, \sigma^2\mathbf{I}_{nl})$ for $\sigma \in \{0, 0.05, 0.1, 0.15\}$ and describe what you see.

2. **Variational reconstruction.** Our goal here is to find a better strategy to reconstruct an image from noisy measurements $\mathbf{y}_\eta$ in a real world setting. For that, we will rewrite the problem to an equivalent problem that is more scalable to solve and we will try to modify this problem to cope with the above issues:

   (a) In practice, l might be very large. So large that even storing the matrix $\mathbf{A}$ is not feasible. So explicitly solving a linear system as done in 1b will not be the way to go (apart from its current issues). Instead we will try to solve adaptations of the minimization problem

   $$\min_{\mathbf{x} \in \mathbb{R}^{l^2}} \|\mathbf{A}\mathbf{x} - \mathbf{y}_\eta\|^2. \tag{3}$$

   Show that $\mathbf{A}^\dagger\mathbf{y}_\eta$ solves the least squares problem (3). It suffices to check first-order necessary conditions.

(b) From 1 we know that solving (3) will not be go down great. Instead we will add *regularization*. That is, we propose to solve the minimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^{l^2}} \|\mathbf{A}\mathbf{x} - \mathbf{y}_\eta\|^2 + \lambda\|\mathbf{x}\|^2, \tag{4}$$

where $\lambda > 0$. If the first term in (4) encodes the information that $\mathbf{x}$ should fit the data, what does the second term – the regularization term – encode?

(c) The first-order optimality conditions for (4) are again a linear system. Write down this system of equations in matrix notation.

(d) The matrix in 2c is symmetric. Compute and plot the eigenvalues this matrix for $n = l/8$, noise η as in 1b with $\sigma^2 = 0.15$ and $\lambda = \{0.1, 0.2, 0.3\}$, and describe what you see (compared to 1b).

3. **Optimization.** Our goal here is to see the effect of the parameter λ on the performance of optimization algorithms. For this, we will implement two methods we have learned in the course:

   (a) Implement gradient descent with fixed step size from $\mathbf{x}^0 = \mathbf{0} \in \mathbb{R}^{l^2}$ for the same settings as in 2d. Report explicitly how you picked the step size (*Hint:* the eigenvalues computed in 1d will come in helpful) and stopping criterion, and plot the progression of the losses $(= \|\mathbf{A}\mathbf{x}^k - \mathbf{y}_\eta\|^2 + \lambda\|\mathbf{x}^k\|^2)$ for all three λ in one plot. Describe what you see.

   (b) Repeat 3a for conjugate gradient.

4. **Model refinement.** Our goal here is to revisit our modeling assumption that led to (4) and try something better, but more tricky to optimize. For that we will try a different regularizer:

   (a) We know that the reconstruction we try to find is *sparse*, i.e., most pixels have value 0. The ℓ^2-regularizer in (4) does not really encode this... There is a better choice to enforce these type of solutions: ℓ^1-regularization. That is, we now aim to solve the minimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^{l^2}} \|\mathbf{A}\mathbf{x} - \mathbf{y}_\eta\|^2 + \lambda\sum_{i=1}^{l^2} |\mathbf{x}_i|. \tag{5}$$

   The absolute value is not differentiable at 0. This could lead to trouble (but it is worth it!). To see some potential issues, minimize the 1D problem $f(t) = |t|$ using gradient descent with fixed step size 1 from $t^0 = 1\frac{3}{4}$. Plot the progression of the loss $(= |t^k|)$ and explain the behaviour you see.

   (b) Since the ℓ^1 regularizer is non-differentiable (even though only at $t = 0$), the theory from class does not hold. We will try making some modifications to gradient descent. Try again to minimize the 1D function in 4a, but now with adaptive step size $\alpha_k = \frac{1}{k+1}$. Again, plot the progression of the loss $(= |t^k|)$.

   (c) Finally, solve (5) with $\lambda = 0.001$ using gradient descent with an adaptive step size similar to 4b (but potentially a different constant than 1 in the numerator) and $\mathbf{x}^0 = \mathbf{0} \in \mathbb{R}^{l^2}$ (*Hint:* the choice for step size in 3b will come in helpful). Report your stopping criterion, plot the progression of the loss and plot the image. Compare the reconstruction to those from 2b.