Deep Learning and Inverse Problems
Summer 2025
Reinhard Heckel

**Problem Set 2 - Chapter 3**
Issued: Thursday May 6, 2025
Due: Thursday May 13, 2025.

---

**Problem 1** (Proximal gradient method for sparse reconstruction)**.** In this problem, we numerically solve a variant of $\ell_1$-minimization for sparse reconstruction. Consider the following $\ell_1$-norm regularized least-squares regression problem:

$$\text{minimize}_{\mathbf{x}} g(\mathbf{x}), \quad g(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1.$$

We take $\mathbf{A} \in \mathbb{R}^{2000\times 1000}$ as a Gaussian random matrix with entries of zero mean and unit variance, and let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, where $\mathbf{e}$ is zero mean Gaussian noise with variance 0.1, and $\mathbf{x}$ is 100-sparse with Gaussian entries, again zero mean and unit variance. Choose the regularization parameter $\lambda$ so that the solution is sparse.

(a) An algorithm to solve this optimization problem numerically is the Iterated Soft Thresholding Algorithm (ISTA), which is an proximal gradient descent algorithm. Define $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$, and $\nabla f(\mathbf{x})$ denotes the gradient of $f$ at $\mathbf{x}$. The ISTA algorithm starts at a random initial point $\mathbf{x}^0$ and then iterates:

$$\mathbf{x}^k = \tau_{\lambda\eta}(\mathbf{x}^k - \eta\nabla f(\mathbf{x}^k)).$$

Here, $\eta$ is a stepsize parameter, and $\tau$ is the soft thresholding operator, which is applied entrywise and is defined as

$$\tau_\lambda(x) = \begin{cases} x + \lambda, & \text{if } x < -\lambda, \\ 0, & \text{if } x \in [-\lambda, \lambda], \\ x - \lambda, & \text{if } x > \lambda. \end{cases}$$

Implement the ISTA algorithm, and provide a plot of convergence of the ISTA algorithm. Plot the objective at iteration $k$ vs the minimal objective, i.e., $g(\mathbf{x}^k) - g(\mathbf{x}^*)$ over the number of iterations, $k$. The minimizer $\mathbf{x}^*$ can be obtained by running the algorithm until convergence.

(b) Next, take $\mathbf{A} \in \mathbb{R}^{500\times 2000}$ as a Gaussian random matrix with zero-mean entries. Generate random $k$-sparse vectors $\mathbf{x}^* \in \mathbb{R}^{2000}$ (as before, take the non-zeros as zero-mean Gaussian) for $k$ ranging from 1 to 500 in some suitably chosen steps, and explore numerically for which values of $k$, recovery of $\mathbf{x}^*$ is possible from the underdetermined measurement $\mathbf{y} = \mathbf{A}\mathbf{x}^*$ by solving the least-squares regression problem numerically. Note that you need to tune the regularization parameter $\lambda$ appropriately. One approach to do so is to perform a grid search of the best parameter $\lambda$ on a logarithmic scale.