Exercises

13.1. **Step Size Rules for Least-Squares Problems**
Consider the gradient method applied to the least-squares objective function $g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2$, i.e.,

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - t_k \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}^{(k)} - \boldsymbol{b}), \quad k = 0, 1, 2 \ldots,$$

where $\boldsymbol{x}^{(0)}$ is an initial guess. For each of the following step size rules, show that the gradient iteration can be implemented such that each iteration only requires a single matrix-vector multiplication with $\boldsymbol{A}$ and one with $\boldsymbol{A}^T$.

1. The step size t_k is constant, i.e., $t_k = t > 0$ for all k.
2. The step size t_k is found by means of the exact line search (13.4).
3. The step size t_k is found by means of a backtracking line search.

13.2. **Lipschitz Continuous Gradients**
Suppose $g_1 \colon \mathbb{R}^n \to \mathbb{R}$ and $g_2 \colon \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable functions. Show that if ∇g_1 and ∇g_2 are Lipschitz continuous with constants L_1 and L_2, respectively, then $\nabla g(\boldsymbol{x}) = \nabla g_1(\boldsymbol{x}) + \nabla g_2(\boldsymbol{x})$ is Lipschitz continuous with constant $L = L_1 + L_2$.

13.3. **SIRT-Like Methods**
Recall that the SIRT iteration (13.15) solves a weighted least-squares problem of the form

$$\text{minimize} \ \ \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_{\boldsymbol{M}}^2 \ ,$$

where the matrix $\boldsymbol{M}$ is symmetric and positive definite.

1. Show that $\|\boldsymbol{M}^{1/2}\boldsymbol{A}\,\boldsymbol{D}^{1/2}\|_2 \leqslant 1$ if $\boldsymbol{M}$ and $\boldsymbol{D}$ are diagonal matrices that satisfy (13.19), i.e.,

$$\boldsymbol{D}_{jj} = \left(\sum_{i=1}^{m} |\boldsymbol{A}_{ij}|^{\alpha}\right)^{-1}, \qquad \boldsymbol{M}_{ii} = \left(\sum_{j=1}^{n} |\boldsymbol{A}_{ij}|^{2-\alpha}\right)^{-1}, \qquad \alpha \in [0, 2]\,.$$

   Hint: Show that $\|\boldsymbol{M}^{1/2}\boldsymbol{A}\,\boldsymbol{D}^{1/2}\boldsymbol{x}\|_2^2 \leqslant \|\boldsymbol{x}\|_2^2$ when $\alpha \in [0, 2]$.
2. Implement the SIRT iteration (13.15) in MATLAB with α as an input parameter.
3. Use your implementation to compute reconstructions for different values of α (say, 0, 1/2, 1, 3/2, and 2). Use the following code to generate a test problem:

```
>> I0 = 1e4;
>> n = 128;
>> A = paralleltomo(n)*(2/n);
>> x = reshape(phantomgallery('grains',n),[],1);
>> I = poissrnd(I0*exp(-A*x));
>> b = -log(I/I0);
```

Compare the reconstructions.

13.4. **Strong Convexity**
Suppose g is a twice continuously differentiable and strongly convex function with strong convexity parameter μ.

1. Show that the smallest eigenvalue of $\nabla^2 g(\boldsymbol{x})$ is bounded by μ.
2. Consider the regularized least-squares objective function for Tikhonov regularization
$$g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \frac{\delta}{2}\|\boldsymbol{x}\|_2^2, \qquad \delta > 0.$$
Derive the Lipschitz constant L associated with the gradient of g and a lower bound on the strong convexity parameter μ.

13.5. **Poisson Measurement Model**
Recall that the negative log-likelihood function associated with the Poisson measurement model may be expressed as
$$g(\boldsymbol{x}) = \mathbf{1}^T \exp(-\boldsymbol{A}\boldsymbol{x}) + \exp(-\boldsymbol{b})^T \boldsymbol{A}\boldsymbol{x} + \text{const.},$$
where $\mathbf{1}$ is the vector of all ones, $\boldsymbol{b} = -\log(\boldsymbol{I}/I_0)$, and the vector $\boldsymbol{I}$ is assumed to be positive.

1. Show that $g(\boldsymbol{x})$ is a convex function of $\boldsymbol{x}$.
2. Derive the first-order optimality condition associated with the ML estimation problem
$$\hat{\boldsymbol{x}}_{\text{ml}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \{g(\boldsymbol{x})\}.$$
3. Show that the gradient of $g(\boldsymbol{x})$ is Lipschitz continuous on $\mathbb{R}^n_+$, i.e., there exists a constant L such that
$$\|\nabla g(\boldsymbol{y}) - g(\boldsymbol{x})\|_2 \leqslant L\|\boldsymbol{y} - \boldsymbol{x}\|_2 \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n_+ .$$
4. Show that if the system of equations $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$ is consistent, then $\boldsymbol{x}$ satisfies the first-order optimality condition $\nabla g(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$.

13.6. **Step Sizes**
In this exercise, we will apply the gradient method to the problem of minimizing
$$g(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2 ,$$

where $\boldsymbol{A}$ and $\boldsymbol{b}$ are generated as follows:

```
>> I0 = 1e6;
>> n = 128;
>> A = paralleltomo(n)*(2/n);
>> x = reshape(phantomgallery('grains',n),[],1);
>> I = poissrnd(I0*exp(-A*x));
>> b = -log(I/I0);
```

Plot the objective value for the first 200 iterations of the gradient method for each of the following step size rules:

1. exact line search (13.7),
2. backtracking line search (Section 13.2.2),
3. BB1 step size (13.39), and
4. BB2 step size (13.40).

Use a semilogarithmic y-axis.

13.7. **Smooth Approximation of the TV Penalty**
Show that the smooth approximations (13.59), (13.61), and (13.63) of the absolute value function all have a Lipschitz continuous derivative with Lipschitz constant $L = 1/\delta$.

13.8. **Regularized Weighted Least-Squares Problems**
Consider the following weighted least-squares problems with two different regularization terms: (i) generalized Tikhonov regularization,

$$\boldsymbol{x}_{\text{GTik}} = \operatorname*{argmin}_{\boldsymbol{x}} \left\{ \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \alpha\,\frac{1}{2}\|\boldsymbol{D}\,\boldsymbol{x}\|_2^2 \right\}, \tag{13.65}$$

and (ii) TV regularization,

$$\boldsymbol{x}_{\text{TV}} = \operatorname*{argmin}_{\boldsymbol{x}} \left\{ \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \alpha\|\boldsymbol{D}\,\boldsymbol{x}\|_1 \right\}, \tag{13.66}$$

where $\boldsymbol{D}$ is defined as in (13.57). The variable $\boldsymbol{x} \in \mathbb{R}^n$ represents an image of size $N \times N$ (i.e., $n = N^2$).

1. Generate a test problem as follows:

```
>> I0 = 1e3;
>> n = 128;
>> A = paralleltomo(n)*(2/n);
>> x = reshape(phantomgallery('grains',n),[],1);
>> I = poissrnd(I0*exp(-A*x));
>> b = -log(I/I0);
```

2. Use power iteration to estimate a Lipschitz constant for the gradient of $g(\boldsymbol{x})$ in the generalized Tikhonov problem (13.65). Plot the estimated Lipschitz constant for different values of γ.

3. Implement and test the PG for solving the minimization problem in (13.65).

4. Implement and test the APG method for solving the minimization problem in (13.65).

5. Implement the APG method for minimizing a smooth approximation of the TV-regularized least-squares problem (13.66), i.e.,

$$\boldsymbol{x}_{\mathrm{TV}} \approx \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\boldsymbol{b} - \boldsymbol{A}\,\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \gamma \sum_{i=1}^{2n} \phi_\delta(\boldsymbol{d}_i^T \boldsymbol{x}) \right\},$$

where $\phi_\delta(\tau)$ is one of the three smooth approximations from Section 13.4.2. Show that the gradient is Lipschitz continuous, and derive a Lipschitz constant.

6. Use your implementations to compute the reconstructions $\boldsymbol{x}_{\mathrm{GTik}}$ and $\boldsymbol{x}_{\mathrm{TV}}$ for different regularization parameters γ. Plot the error norms $\|\overline{\boldsymbol{x}} - \boldsymbol{x}_{\mathrm{TV}}\|_2$ and $\|\overline{\boldsymbol{x}} - \boldsymbol{x}_{\mathrm{GTik}}\|_2$ versus γ. Compare the “best” reconstructions from both models.