18.327 Computational Inverse Problems – Spring 2018
Problem set 3 – Due 04/12/2018

Problems are labeled ($\star$) for easy, ($\star\star$) for medium, and ($\star\star\star$) for hard. For homework 3, solve (at least) five stars worth of questions. Recommended exercises: 1, 2, 5, 6.

1. ($\star$) Let $y_i = x + e_i$ for $i = 1, \ldots, n$, with $x \in \mathbb{R}$, and $e_i \sim N(0, \sigma^2)$ i.i.d. The MLE for x is the sample mean (empirical average) $\hat{x} = \frac{1}{n} \sum_i y_i$. In the frequentist framework, x is fixed, and $\hat{x}$ has a distribution $p(\hat{x}|x)$. In the Bayesian framework with a uniform prior, x is random, and has a distribution $p(x|y)$. Even though $p(\hat{x}|x)$ and $p(x|y)$ are philosophically different, show that their expressions coincide in this particular example. Generalize your argument to a general linear model of the form $y_i = a_i^T x + e_i$, where $x \in \mathbb{R}^m$.

2. ($\star$) *(Poisson noise)* Suppose that a gamma-ray detector counts photons in n energy bins, yielding integer measurements y_i, with $i = 1, \ldots, n$. The photon counts y_i are independent and modeled by Poisson distributions $P(y_i = x) = \frac{\lambda_i^x e^{-\lambda_i}}{x!}$, but the intensities λ_i that give rise to those counts are assumed to follow a power law $\lambda_i = \mu \epsilon_i$, for some unknown μ, but known numbers ϵ_i in geometric progression. Find the maximum likelihood estimator (MLE) of μ. [Hint: assume that $\lambda_i = \mu \epsilon_i$ goes in the forward model, not in the prior.]

3. ($\star\star$) In the setting of the previous question, find or characterize the maximum likelihood estimator of λ_i and μ, when a relaxed form of $\lambda_i = \mu \epsilon_i$ is imposed via the prior $p(\lambda_1, \ldots, \lambda_n, \mu) \sim \exp\left[-\frac{1}{2\delta^2} \sum_i (\lambda_i - \mu \epsilon_i)^2\right]$, for some small $\delta > 0$, rather than through the forward model.

4. ($\star$) *Jeffreys parameters, continued.* In homework 1, we encountered positive parameters, whose distance is better measured via $|\log \sigma_1 - \log \sigma_2|$ than via $|\sigma_1 - \sigma_2|$. Consider a Bayesian prior $p(\sigma)$ which is objective for such (Jeffreys) parameters, in the sense that it is a (unnormalized) uniform probability distribution for $\log \sigma$.

   (a) What prior distribution does this give rise to, for σ itself?

   (b) What prior distribution does this give rise to, for σ^a with $a > 0$?

5. ($\star$) Consider $y = Ax_0 + e$ with $e_i \sim N(0, \sigma^2)$ i.i.d., and the Tykhonov-regularized least-squares problem
$$\min \|Ax - y\|_2^2 + \lambda^2 \|x\|_2^2.$$
   In this exercise, suppose that σ and $\|x_0\|$ are known, and assume that A is square and a multiple of an isometry i.e., $A^T A = AA^T = a^2 I$. Find λ for which the MSE $\mathbb{E}\|x - x_0\|^2$ is minimum. (The solution to this exercise illustrates the following useful heuristic: λ should be chosen so that the misfit and regularization terms are of comparable size.)

6. ($\star\star$) *Regularization by terminating the iterations.* Consider solving $Ax = y$ with square A, by gradient descent:
$$x_{n+1} = x_n + \alpha A^T (y - Ax), \qquad x_0 = 0,$$
   for sufficiently small $\alpha > 0$. Let $A = U \Sigma V^T$ be the singular value decomposition of A; at convergence,
$$x_\infty = \sum_i v_i \sigma_i^{-1} u_i^T y.$$

   (a) Find a function $f_n(\sigma)$ for which
$$x_n = \sum_i v_i f_n(\sigma_i) \sigma_i^{-1} u_i^T y.$$

   (b) In the limit of large n, show that x_n is close to the solution of a Tykhonov-regularized least-squares problem, i.e., find $\lambda(n)$ for which x_n is well-approximated by the solution of
$$\min \|Ax - y\|_2^2 + \lambda(n)^2 \|x\|_2^2.$$

   (In practice, coupling the formula that you obtain for $\lambda(n)$, with the conclusion of the previous exercise, shows a useful way to choose n for inverse problems.)