Midterm exam
Friday, March 8, 2019

SOLUTIONS

1 (80 Points) Short Questions

A. (10 Points) Suppose you are trying to minimize a cost function $f(\mathbf{x})$, and you find a point $\mathbf{x}^{(0)}$ where the gradient is zero. For an arbitrary (not necessarily convex) cost function, is this point guaranteed to be a global minimizer to the cost function $f(\mathbf{x})$? In other words, can we guarantee that the cost function $f(\mathbf{x}^{(0)}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^N$? Why or why not?

Answer: No - no guarantees of global optimality. If the function is non-convex, in general a location with gradient zero may correspond to a local minimizer, a saddle point, or a local maximizer. So, no guarantees of local optimality, let alone global optimality.

B. (10 Points) In general, steepest descent converges in fewer iterations than conjugate gradients and Newton's method (TRUE/FALSE)

Answer: FALSE. Steepest descent generally takes more iterations than conjugate gradients or Newton's method (although the specifics depend on the cost function and initialization).

C. (10 Points) If we acquire Fourier data, where we can model the data as a DFT (in 2D or 3D) of the true image, we often reconstruct the image by doing a simple inverse DFT (in 2D or 3D, as appropriate). (TRUE/FALSE)

Answer: TRUE. DFT is the workhorse of Fourier imaging techniques.

D. (5 Points) If we acquire projection data, where we can model the data as a set of projections of our image along different angles, we often reconstruct our image through (unfiltered) back-projection of these projection data (TRUE/FALSE).

Answer: FALSE. Unfiltered backprojection will give rise to a heavily smoothed version of our image.

E. (5 Points) If we acquire projection data, where we can model the data as a set of projections of our image along different angles, we often reconstruct our image through filtered backprojection (FBP) of these projection data (TRUE/FALSE).

Answer: TRUE. FBP will result in good image reconstruction if we have sufficient projection data.

F. (10 Points) Say we have an image reconstruction formulation that includes a term of the form $\mathbf{Ax}$, where $\mathbf{x}$ is our entire image (of size $N \times N$) reshaped as a vector of length N^2. Then, in order to solve our formulation we always need to explicitly create the matrix $\mathbf{A}$ (which can be enormous, eg: $N^2 \times N^2$). For instance, if the matrix performs a DFT in 2D, we will need to explicitly create each entry of the matrix in a huge array in order to solve this formulation with one of our descent algorithms. (TRUE/FALSE)

Answer: FALSE. As illustrated in our homework sets and lectures, we do not need to create the full matrix $\mathbf{A}$ in order to use it in reconstruction, for a lot of useful operations including DFTs, projections, filtering, subsampling, etc.

G. (10 Points) Say we are trying to reconstruct an image like we did in homework set 3. Since Newton's algorithm is quite powerful, why not just use this algorithm (as we did in homework 2) by creating the corresponding Hessian matrix and inverting it to obtain $\mathbf{x}_1 = \mathbf{x}_0 - \mathbf{H}^{-1}\mathbf{g}$? What is the main challenge for this approach when reconstructing images (or large vectors)?

Answer: The main challenge is computational: explicitly creating and inverting the full Hessian matrix is enormously costly (in terms of computation) for large image reconstruction problems.

H. (10 Points) Suppose we have two reconstruction problems where we use an l2-regularized LS reconstruction (the same for both problems):

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{d}\|_2^2 + \lambda \|\mathbf{Cx}\|_2^2 \tag{1.1}$$

For the first, the data are $\mathbf{d} = \mathbf{d}_1$, and for the second, the data are $\mathbf{d} = \mathbf{d}_2$. In other words, our two image reconstruction problems are identical except for the data used in each of them. These problems have solutions $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$, respectively. Now suppose we have a third problem, with the same reconstruction formulation but this time the data are $\mathbf{d}_3 = \mathbf{d}_2 - \mathbf{d}_1$. Can you express the solution $\hat{\mathbf{x}}_3$ in terms of $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$? Please provide the corresponding (simple) expression and explain how you derived it.

Answer: The solution to our third problem will be:

$$\hat{\mathbf{x}}_3 = \hat{\mathbf{x}}_2 - \hat{\mathbf{x}}_1$$

given the linear nature of the solution to a linear least-squares problem.

I. (10 Points) If we pose our image reconstruction using an ℓ_p regularized formulation:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{d}\|_2^2 + \lambda \|\mathbf{Cx}\|_p^p \tag{1.2}$$

for some $p \geq 1$, the specific choice of p won't matter much, ie: the reconstructed image will look identical regardless of the value of p. (TRUE/FALSE)

Answer: FALSE. The reconstructed image is highly dependent on the choice of regularization (eg: $p = 1$ vs $p = 2$), as we have discussed in class. For instance, ℓ_2 regularized images are typically much smoother than ℓ_1 regularized images.

2 (20 Points) Constrained Reconstruction

Suppose we are reconstructing an image $\mathbf{x}$ from some data $\mathbf{d} = \mathbf{Ax}$. Further, suppose we know a priori which pixels within the image correspond to tissues (a total of N_T pixels where the image intensity is different from zero), and which pixels correspond to air (a total of N_A pixels where the image intensity is zero). The total number of pixels in the image is $N = N_T + N_A$. This constraint can be expressed as $\mathbf{Cx} = \mathbf{0}$, where $\mathbf{C}$ is a subsampling matrix that selects all the pixels in the air regions (and ignores the rest), and $\mathbf{0}$ is a zero-vector of the appropriate length N_A.

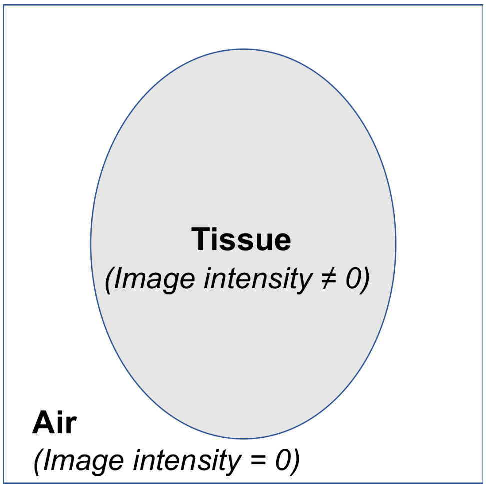

Figure 2.1: Image reconstruction in a case where we know where the air (with image intensity zero) is located.

Putting all these elements together, we have the following constrained formulation:

$$\begin{aligned} \hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{d}\|_2^2 \\ \text{such that } \mathbf{Cx} = \mathbf{0} \end{aligned} \tag{2.1}$$

2.1 (5 points) Is our optimization problem convex?

Explain why or why not based on the cost function and the constraint. Can we guarantee a globally optimal solution by a descent algorithm that is able to find a locally optimal solution?

Answer: Yes, it is convex. Our cost function is a sum of squares ($\|\cdot\|_2^2$, which is convex), inside the sum of squares we have an affine function, so it stays convex, and our constraint is a linear function, which is also convex. Therefore, the overall constrained optimization problem is convex and a descent algorithm that converges to a locally optimal solution should also be globally optimal.

2.2 (10 points) Rewrite as unconstrained problem

Rewrite this constrained optimization problem (equation 2.1 above) as an unconstrained optimization problem, solving for a vector $\mathbf{y}$ of length N_T (ie: the intensities at all the tissue pixels, since we know the remaining N_A pixels are zero).

Write this problem as an unconstrained problem in terms of $\mathbf{A}$, $\mathbf{d}$, and $\mathbf{C}$.

Answer: We can rewrite our entire image $\mathbf{x}$ as $\mathbf{x} = \mathbf{D}^H\mathbf{y}$ where $\mathbf{y}$ is a vector that includes the image intensities of $\mathbf{x}$ over the tissue voxels (ie: $\mathbf{y}$ is a vector of length N_T), and $\mathbf{D}^H$ is a zero-padding operation that takes $\mathbf{y}$ and pads it with zeros in the air regions of the image. Note that $\mathbf{D}^H$ also happens to be the transpose of the matrix $\mathbf{D}$, which subsamples the tissue region instead of the air region as done by $\mathbf{C}$. In other words, $\mathbf{D}$ is the subsampling matrix such that:

$$\mathbf{D}^H\mathbf{D} = \mathbf{I} - \mathbf{C}^H\mathbf{C}$$

where $\mathbf{I}$ is the identity matrix.

Importantly, expressing our image as $\mathbf{x} = \mathbf{D}^H\mathbf{y}$ inherently includes the constraint that the image $\mathbf{x}$ should be zero over the air regions (ie: $\mathbf{C}\mathbf{D}^H\mathbf{y} = 0$ by construction, for any choice of $\mathbf{y}$).

With this modification, we can rewrite our problem as the following unconstrained problem:

$$\hat{\mathbf{y}} = \arg\min_{\mathbf{y}} \|\mathbf{A}\mathbf{D}^H\mathbf{y} - \mathbf{d}\|_2^2$$

which, as described in this course, has closed form solution:

$$\hat{\mathbf{y}} = \left[(\mathbf{A}\mathbf{D}^H)^H\mathbf{A}\mathbf{D}^H\right]^{-1}(\mathbf{A}\mathbf{D}^H)^H\mathbf{d}$$

Partial credit will be given for alternative solutions, eg: where $\lambda\|\mathbf{C}\mathbf{x}\|_2^2$ is added to the cost function, leading to a *related* (but generally not identical) unconstrained optimization problem.

2.3 (5 points) Closed form solution

Provide a closed form solution for $\hat{\mathbf{x}}$ in terms of $\mathbf{A}$, $\mathbf{d}$, and $\mathbf{C}$.

Answer: Since our image $\mathbf{x}$ is simply $\mathbf{x} = \mathbf{D}^H\mathbf{y}$ (and $\mathbf{D}^H$ can be easily derived if we have knowledge of $\mathbf{C}$ as described above), we can obtain our entire solution image $\hat{\mathbf{x}}$ as:

$$\hat{\mathbf{x}} = \mathbf{D}^H\left[(\mathbf{A}\mathbf{D}^H)^H\mathbf{A}\mathbf{D}^H\right]^{-1}(\mathbf{A}\mathbf{D}^H)^H\mathbf{d} = \mathbf{D}^H\left[\mathbf{D}\mathbf{A}^H\mathbf{A}\mathbf{D}^H\right]^{-1}\mathbf{D}\mathbf{A}^H\mathbf{d}$$

Partial credit will be given for alternative solutions, eg: solving $\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{d}\|_2^2 + \lambda\|\mathbf{C}\mathbf{x}\|_2^2$ which is a related optimization problem to our original problem (although generally not identical).