

Lecture 37

Recap of Optimization in Imaging and Therapy

37.1 Lecture Objectives

- Recognize that the same underlying optimization concepts apply to both imaging and therapy scenarios, as well as various other applications.
- Identify convex and non-convex optimization problems.
- Recognize easy optimization problems (particularly linear least squares) and the many scenarios where they arise.
- Recognize specific challenges in more difficult problems, and develop awareness of strategies to diagnose or address these challenges.

37.2 Introduction

Optimization problems come up in a variety of medical physics and biomedical applications. Indeed, they come up in various scientific and technological settings as well as everyday life scenarios (although, admittedly, we do not always stop to write down an explicit cost function). This lecture will review and explore additional applications for some of the concepts covered in our course, to illustrate the broad applicability of these concepts.

37.3 Mapping practical problems to optimization problems

In the past few lectures, we have reviewed several applications of optimization in imaging and in therapy. In each application, we typically seek to determine some parameters that

best fit some acquired data as well as some prior knowledge, or we seek to determine some experimental parameters that best accomplish some goal, etc. Many problems across all sorts of applications map nicely to least squares regression, whether linear or non-linear. Thus, we will briefly review least-squares regression next.

37.3.1 Linear least-squares regression

Linear least-squares (LLS) problems were described in some detail earlier in this course. The LLS problem can be written as follows:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \quad (37.1)$$

LLS problems have a simple, closed-form solution, as follows:

$$\mathbf{x}^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \quad (37.2)$$

which we can apply as long as the matrix $\mathbf{A}^T \mathbf{A}$ has an inverse matrix $(\mathbf{A}^T \mathbf{A})^{-1}$ ¹

Recall the broad utility of this formulation and solution, which can conceivably be used in a variety of settings, simply by adjusting the interpretation (and dimensions, and values) of $\mathbf{x}$, $\mathbf{A}$, and $\mathbf{y}$. For example, in image reconstruction from linear measurements (such as Fourier transforms or projections), $\mathbf{x}$ is the reconstructed image, $\mathbf{y}$ represents the acquired data in the corresponding domain, and $\mathbf{A}$ represents the imaging physics and acquisition parameters, which map the desired image to the corresponding measurements. Alternatively, in therapy planning, $\mathbf{x}$ may denote the fluence map we would like to design, $\mathbf{y}$ may represent the desired dose deposition map, and $\mathbf{A}$ is the dose deposition matrix. In this lecture, we present a different interpretation of LLS problems in the context of statistical estimation. Since we have not covered probability in this course, this interpretation is presented as an interesting data point (or piece of “trivia”), but will not be thoroughly treated or required for the final exam.

37.3.2 Example: Maximum-Likelihood Estimation

Let us formulate a Maximum-Likelihood (ML) estimation problem in the presence of Gaussian noise as a LS optimization problem. Suppose we have a measuring device (eg: an imaging system) that takes measurements from an unknown object determined by a vector $\mathbf{x}$, as follows:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \eta \quad (37.3)$$

where $\mathbf{b} \in \mathbb{R}^M$ is our measurement vector, $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the system matrix, $\mathbf{x} \in \mathbb{R}^N$ is the unknown vector that we would like to estimate, and $\eta \in \mathbb{R}^M$ is a noise vector containing independent, identically distributed (iid) noise in each of its entries. Further, assume that the noise distribution is Gaussian with mean zero and standard deviation σ_1 .

¹Think about the implications of the matrix $\mathbf{A}^T \mathbf{A}$ not being invertible. For what kinds of matrices $\mathbf{A}$ will this happen? How can we deal with this situation?

Question: Derive the maximum-likelihood (ML) estimate for $\mathbf{x}$ as a least-squares (LS) optimization problem, and describe whether this optimization problem depends on σ (ie: will the ML solution $\mathbf{x}_{ML}^*$ change based upon the noise level?).

Hint: Note that the ML estimate $\mathbf{x}_{ML}^*$ corresponds to maximizing the probability of observing $\mathbf{b}$ given $\mathbf{x}$, ie:

$$\mathbf{x}_{ML}^* = \arg \max_{\mathbf{x}} P(\mathbf{b}|\mathbf{x}) \quad (37.4)$$

where, in the presence of Gaussian noise, $P(\mathbf{b}|\mathbf{x})$ is given by:

$$P(\mathbf{b}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^M \sigma_1^{2M}}} e^{-\frac{\|\mathbf{Ax}-\mathbf{b}\|_2^2}{2\sigma_1^2}} \quad (37.5)$$

SOLUTION:

$$\begin{aligned} \mathbf{x}_{ML}^* &= \arg \max_{\mathbf{x}} P(\mathbf{b}|\mathbf{x}) \\ &= \arg \max_{\mathbf{x}} \frac{1}{\sqrt{(2\pi)^M \sigma_1^{2M}}} e^{-\frac{\|\mathbf{Ax}-\mathbf{b}\|_2^2}{2\sigma_1^2}} \\ &= \arg \max_{\mathbf{x}} \log \left[\frac{1}{\sqrt{(2\pi)^M \sigma_1^{2M}}} e^{-\frac{\|\mathbf{Ax}-\mathbf{b}\|_2^2}{2\sigma_1^2}} \right] \\ &= \arg \max_{\mathbf{x}} \left[\log \left(\frac{1}{\sqrt{(2\pi)^M \sigma_1^{2M}}} \right) - \frac{\|\mathbf{Ax} - \mathbf{b}\|_2^2}{2\sigma_1^2} \right] \\ &= \arg \max_{\mathbf{x}} [-\|\mathbf{Ax} - \mathbf{b}\|_2^2] \\ &= \arg \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2 \end{aligned}$$

where we can apply a log transformation to our objective function, since the log function is monotonically increasing for arguments greater than zero, so it will preserve the location of the minimizer. We can also remove constant offset terms ($\log \left(\frac{1}{\sqrt{(2\pi)^M \sigma_1^{2M}}} \right)$) from the maximization since they do not depend on $\mathbf{x}$. Subsequently, we can remove constant multiplicative terms ($\frac{1}{2\sigma_1^2}$) since they do not depend on $\mathbf{x}$ either. Finally, maximizing a function $-f(x)$ is equivalent to minimizing $f(x)$, as applied in the last step.

Importantly, the approach to obtain an ML solution to this estimation problem is independent of the noise level. In the presence of iid Gaussian noise, ML is equivalent to LS fitting, ie: minimizing the ℓ_2 norm (or ℓ_2 norm squared) of the data fitting residual $\mathbf{Ax} - \mathbf{b}$.

37.3.3 Maximum-a-Posteriori (MAP) Estimation

Now let us assume that we know the a priori statistical distribution of $\mathbf{x}$, which happens to be a Gaussian iid distribution with standard deviation σ_2 . In other words, in the absence of any data, the probability density function of $\mathbf{x}$ is given by:

$$P(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \sigma_2^{2N}}} e^{-\frac{\|\mathbf{x}\|_2^2}{2\sigma_2^2}} \quad (37.6)$$

Now that we have both an a priori distribution as well as some noisy data, let's formulate the maximum-a-posteriori (MAP) estimation of $\mathbf{x}$ (which combines both sources of information), as follows:

$$\mathbf{x}_{MAP}^* = \arg \max_{\mathbf{x}} P(\mathbf{x}|\mathbf{b}) \quad (37.7)$$

$$= \arg \max_{\mathbf{x}} \frac{P(\mathbf{b}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{b})} \quad (37.8)$$

$$= \arg \max_{\mathbf{x}} P(\mathbf{b}|\mathbf{x})P(\mathbf{x}) \quad (37.9)$$

where the last step is warranted since $P(\mathbf{b})$ does not depend on $\mathbf{x}$.

Question: Formulate the MAP estimation above as a LS optimization problem of the form:

$$\mathbf{x}_{MAP}^* = \arg \min_{\mathbf{x}} \|\mathbf{E}\mathbf{x} - \mathbf{f}\|_2^2 + \lambda \|\mathbf{G}\mathbf{x}\|_2^2 \quad (37.10)$$

and determine the matrices, vectors, and scalar in this LS optimization ($\mathbf{E}$, $\mathbf{f}$, λ , $\mathbf{G}$), in terms of the parameters of our problem ($\mathbf{A}$, $\mathbf{b}$, σ_1 , σ_2 , etc).

SOLUTION:

$$\begin{aligned} \mathbf{x}_{MAP}^* &= \arg \max_{\mathbf{x}} P(\mathbf{b}|\mathbf{x})P(\mathbf{x}) \\ &= \arg \max_{\mathbf{x}} \frac{1}{\sqrt{(2\pi)^M \sigma_1^{2M}}} e^{-\frac{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2}{2\sigma_1^2}} \frac{1}{\sqrt{(2\pi)^N \sigma_2^{2N}}} e^{-\frac{\|\mathbf{x}\|_2^2}{2\sigma_2^2}} \\ &= \arg \max_{\mathbf{x}} \log \left[\frac{1}{\sqrt{(2\pi)^M \sigma_1^{2M}}} e^{-\frac{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2}{2\sigma_1^2}} \frac{1}{\sqrt{(2\pi)^N \sigma_2^{2N}}} e^{-\frac{\|\mathbf{x}\|_2^2}{2\sigma_2^2}} \right] \\ &= \arg \max_{\mathbf{x}} \left[\log \left(\frac{1}{\sqrt{(2\pi)^M \sigma_1^{2M}}} \right) - \frac{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2}{2\sigma_1^2} + \log \left(\frac{1}{\sqrt{(2\pi)^N \sigma_2^{2N}}} \right) - \frac{\|\mathbf{x}\|_2^2}{2\sigma_2^2} \right] \\ &= \arg \max_{\mathbf{x}} \left[-\frac{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2}{2\sigma_1^2} - \frac{\|\mathbf{x}\|_2^2}{2\sigma_2^2} \right] \\ &= \arg \max_{\mathbf{x}} \left[-\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 - \frac{2\sigma_1^2}{2\sigma_2^2} \|\mathbf{x}\|_2^2 \right] \end{aligned}$$

$$\begin{aligned}
&= \arg \min_{\mathbf{x}} \left[\|\mathbf{Ax} - \mathbf{b}\|_2^2 + \frac{\sigma_1^2}{\sigma_2^2} \|\mathbf{x}\|_2^2 \right] \\
&= \arg \min_{\mathbf{x}} \left[\|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_2^2 \right]
\end{aligned}$$

which is a minimization problem with the desired form, with $\mathbf{E} = \mathbf{A}$, $\mathbf{f} = \mathbf{b}$, $\mathbf{G} = \mathbf{I}$ (the identity matrix), and $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$. Importantly, the parameter λ is the ratio of the data noise variance to the variance of the prior distribution of $\mathbf{x}$. This parameter controls the tradeoff between how much we trust our data (if σ_1^2/σ_2^2 is small, then the optimization problem is dominated by the data term $\|\mathbf{Ax} - \mathbf{b}\|_2^2$) and how much we trust our prior distribution (if σ_1^2/σ_2^2 is large, then the optimization problem is dominated by the ‘prior’ (or ‘regularization’) term $\|\mathbf{x}\|_2^2$).

37.3.4 Non-Linear Least-Squares Regression

Similar to the LLS example above, we are often faced with scenarios where we wish to fit a non-linear signal model ($\mathbf{f}(\mathbf{x})$) to some acquired data ($\mathbf{y}$). Based on justifications like ML estimation in the presence of Gaussian noise, this fitting is often performed in the LS sense, leading to the Non-Linear Least-Squares (NLLS) problem:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|_2^2 \quad (37.11)$$

where the vector-valued function $\mathbf{f}(\mathbf{x})$ describes the corresponding non-linear signal model for each of the acquired data points. For example, signals acquired in medical imaging can sometimes be modeled as a decaying exponential of the form:

$$f_n(\mathbf{x}) = x_1 e^{-b_n x_2}, \text{ for } n = 1, 2, \dots, N \quad (37.12)$$

where N data points are measured, each with a certain value of b_n (which can be controlled experimentally), and the signal can be parameterized in terms of two unknown parameters: an initial amplitude (x_1), and an exponential decay rate (x_2). Based on multiple ($N \geq 2$) acquired data points, we are able to estimate these two parameters x_1 and x_2 using NLLS. Figure 37.1 provides an example of NLLS, for R_2^* in Magnetic Resonance Imaging, where the signal can be described as a decaying exponential as a function of increasing echo time (TE): $s(TE_n) = Ae^{-TE_n \cdot R_2^*}$, for $n = 1, \dots, N$, in the presence of additive noise. In this case, the independent variable b_n corresponds to the acquired echo times (TE_n), x_1 corresponds to the initial signal amplitude at $TE = 0$ ms, and x_2 corresponds to the parameter R_2^* , a parameter that describes the rate of decay of MR signals after excitation. The problem of R_2^* mapping is important in MRI, for applications such as iron quantification.

Question: We know that the LLS problem is convex. Is the NLLS problem convex too? How would one prove or disprove this convexity?

Question: How could one solve an NLLS problem such as the example given in this section (with an exponential signal model) in practice?

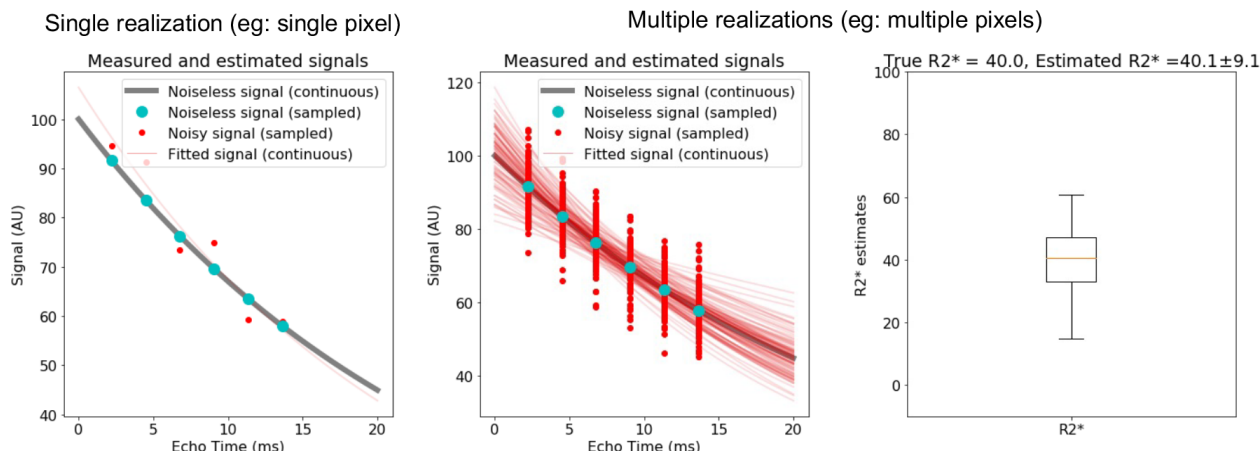


Figure 37.1: Illustrative example of Non-Linear Least Squares (NLLS) fitting as applied to parametric mapping of R_2^* in Magnetic Resonance Imaging, where the signal can be described as a decaying exponential as a function of increasing echo time (TE): $s(TE) = Ae^{-TE \cdot R_2^*}$, in the presence of additive noise. In this experiment, the choice of TE s can be controlled by the user (in this plot, $N = 6$ echo times are acquired). The resulting R_2^* parameter is of high relevance in the clinic as well as in research, as it correlates closely with iron concentration in tissue.

37.3.5 Non-Convexity in NLLS

The lack of convexity of most NLLS problems can have important practical consequences. MRI-based multi-echo acquisitions (such as those used for R_2^* mapping, as described above) are also often used for separation of water and fat signals in tissue. This method is based on the different resonance frequencies of water and fat protons, and is widely used in the clinic as well as in research. However, the problem of water-fat signal separation in MRI is notoriously non-convex, with two main local optimizers: one water-dominated and one fat-dominated. For example, a voxel where fat comprises approximately 20% of the signal will lead to a correct solution (fat fraction of 20%) as well as a ‘swapped’ solution (fat fraction of around 80%). When the wrong solution is reached at some or all the voxels, images of non-diagnostic quality are obtained. This is a major challenge in MRI, which has been the focus of multiple PhD dissertations (including this course’s instructor’s [\[2\]](#)). See figure [37.2](#) for an illustration of this non-convexity induced challenge.

37.4 Volumetric Modulated Arc Therapy

Volumetric Modulated Arc Therapy (VMAT) is a therapy technology where radiation is applied continuously while the gantry rotates (as opposed to IMRT, as discussed previously). In VMAT, the therapy planning (optimization) problem consists of determining

²See, for example: Hernando D, et al, Robust water/fat separation in the presence of large field inhomogeneities using a graph cut algorithm, Magnetic Resonance in Medicine 2010, 63(1):79-90

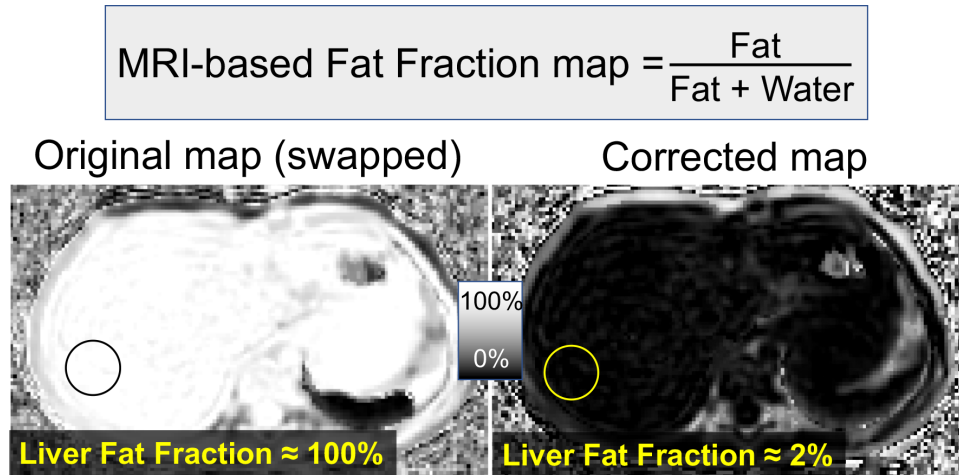


Figure 37.2: Illustration of the non-convexity of water-fat separation in multi-echo MRI, which can lead to erroneous “swapped” solutions where water-dominant voxels are estimated to be fat-dominant, and fat-dominant voxels are estimated to be water-dominant.

the following parameters (each of them as a function of time): i) the MLC leaf trajectories, ii) gantry angle, and iii) dose rate. Direct Aperture Optimization (DAO) are powerful for designing optimized VMAT therapy plans. However, DAO for VMAT is a highly non-convex, mixed-integer optimization problem. In order to address this challenging problem, methods that seek the global solution to the corresponding non-convex optimization have been proposed (eg: using simulated annealing, as in Otto K 2008. Volumetric modulated arc therapy: IMRT in a single gantry arc Med. Phys 35 3107). Alternatively, the non-convex problem has also been formulated as a sequence of convex problems.

37.5 Summary

Optimization problems arise in a variety of scenarios in biomedicine. Of note, linear least-squares (LLS) problems deserve special attention because of their high prevalence, efficient solutions, and well-understood properties. In addition, understanding foundational optimization concepts, such as convexity, allows us to better understand existing solutions to important problems (e.g., in imaging and in therapy), and often also develop better solutions. A combination of theoretical understanding and computational tools, as introduced in this course, should hopefully serve the student well in their academic and professional endeavors.

