MATH 262/CME 372: Applied Fourier Analysis and Elements of Modern Signal Processing Winter 2016

Lecture 10 — February 4, 2016

Prof. Emmanuel Candes

Scribe: Carlos A. Sing-Long, Edited by E. Bates

1 Outline

Agenda: Ill-posed inverse problems

1. Ill-posedness of deconvolution
2. Ill-posedness of the Radon inversion
3. Regularization

Last Time: In modeling the problem of X-ray tomography, we introduced the Radon transform, which outputs a collection of line integrals of an input function. The procedure of backprojection was a first attempt at recovering a signal from its Radon transform; however, we saw that it did not return exact reconstruction, but rather a smoothed version of the original signal. Through the projection-slice theorem, we established a relation between the Radon and the Fourier transforms. This relation allowed us to derive a closed-form expression for the actual inverse Radon transform, which we call the *filtered* backprojection formula.

2 Ill-posedness of the inverse problem

Recall from Lecture 9 the Radon inversion formula, or filtered backprojection:

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^2}\int_0^{\pi} (\mathcal{R}f(\,\cdot\,,\theta) * h)(\langle \boldsymbol{x}, \boldsymbol{n}_\theta\rangle)\, d\theta,$$

where $\boldsymbol{n}_\theta = (\cos\theta, \sin\theta)$, and h is such that $\hat{h}(\omega) = |\omega|$. The inverse transform thus filters the Radon transform by h, which is declared to have Fourier transform $\hat{h}(\omega) = |\omega|$, and *then* performs a scaled backprojection. By analogy with differentiation, for which $\widehat{\frac{d}{dt} f}(\omega) = i\omega \hat{f}(\omega)$, we see that the filter performs an operation similar to a derivative. Roughly speaking, then, the filter makes objects *more* singular. Another perspective is that while the backprojection formula produces images that are blurred, meaning edges are smoothed, convolving with h somehow recreates the edges from smooth data. Indeed, we are making objects more singular.

This feature of Radon inversion raises serious obstacles in its implementation. Consider the unavoidable fact that real-world measurements of a signal $\mathcal{R}f$ will contain noise. And any realization of noise (white, Brownian, grey, pink, etc.) is non-differentiable, meaning a filter that essentially takes derivatives will return a signal that has infinite energy! In a very dramatic way, the Radon inversion *amplifies* noise.

This complication is a result a mathematical observation that the inverse Radon transform

$$\mathcal{R}^{-1} : L^2([0,\pi) \times \mathbb{R}) \to L^2(\mathbb{R}^2)$$

is an unbounded operator, a fact we will see later. Consequently, two images that are far apart may have Radon transforms that are very close. In the presence of even a trivial amount of noise, there is a tremendous amount of ambiguity as to what the actual image is. In fact, it is possible to construct an analytic function whose intergral along *any* line vanishes, and therefore its Radon transform is identically zero, yet the function itself is nonzero. Moreover, one can construct a sequence of functions diverging everywhere but whose Radon transforms converge uniformly (see [5]).

For these reasons, we see that the Radon inverse problem is **ill-posed**. Roughly speaking, an inverse problem is ill-posed when the solution is discontinuous (i.e. highly sensitive) with respect to the final conditions (the observed data). Ill-posed problems arise in many applications, and Radon inversion is just one example. The survey paper [6] provides a wealth of other examples. In the next section we discuss a prototypical case of ill-posedness: deconvolution.

3 Ill-posedness of deconvolution

Suppose we observe a function $g = f * \varphi$, where

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \Rightarrow \quad \hat{\varphi}(\omega) = e^{-\frac{\omega^2}{2}}.$$

In theory, we can recover f from g by direct inversion, since

$$\hat{f}(\omega) = \hat{g}(\omega) e^{\frac{\omega^2}{2}}, \tag{1}$$

and we have a Fourier inversion formula. In reality, however, the observation will come with noise[1], which we denote by z:

$$g = f * \varphi + z \quad \Rightarrow \quad \hat{g}(\omega) = \hat{f}(\omega) e^{-\frac{\omega^2}{2}} + \hat{z}(\omega) \quad \Rightarrow \quad \hat{g}(\omega) e^{\frac{\omega^2}{2}} = \hat{f}(\omega) + \hat{z}(\omega) e^{\frac{\omega^2}{2}}. \tag{2}$$

If we naively use $\hat{g}(\omega)e^{\omega^2/2}$ from (1) as our estimate for $\hat{f}(\omega)$, then the error term $\hat{z}(\omega)e^{\omega^2/2}$ will ensure that we are grossly inaccurately. Indeed, the observation error is amplified *exponentially* in the size of the frequency. Moreover, in the case of white noise, $\hat{z}(\omega)$ does not even decay as $|\omega| \to \infty$, thereby reinforcing the artificial high frequency content introduced to the signal. In any case, the reconstruction coming from (1) will bear no resemblance to the true signal.

4 Ill-posedness of Radon inversion

Computerized tomography is no exception to the presence of noise. Even if one attempted to decrease noise levels to negligible amounts, two problems persist:

[1] Here we work with the common model of noise being additive, although one can also consider multiplicative noise.

1. At some point, engineering an increased signal-to-noise ratio requires a greater flux of radiation. In the case of X-rays, this quite literally burns the human subject and causes genetic mutations. So for obvious health concerns we must settle with a certain level of noise.

2. Since data cannot be measured continuously or with infinite precision, the procedure of discretization introduces truncation errors and gaps in data. Thus, even without physical noise, inaccuracies are inherent in application.

At this point we should be convinced that ill-posedness is unavoidable in Radon inversion. In order to study the nature of this poor conditioning, we first compute the adjoint $\mathcal{R}^*$ of the Radon transform. Formally, we have

$$\begin{aligned} \langle g,\, \mathcal{R}f\rangle &= \int_0^\pi \int_{\mathbb{R}} \overline{g(t,\theta)}\mathcal{R}f(t,\theta)dtd\theta \\ &= \int_0^\pi \int_{\mathbb{R}} \iint_{\mathbb{R}^2} \overline{g(t,\theta)} f(\boldsymbol{x})\delta(\langle \boldsymbol{x},\, \boldsymbol{n}_\theta\rangle - t)\, d\boldsymbol{x}dtd\theta \\ &= \iint_{\mathbb{R}^2} \overline{\left(\int_0^\pi g(\langle \boldsymbol{x},\, \boldsymbol{n}_\theta\rangle, \theta)d\theta\right)} f(\boldsymbol{x})\, d\boldsymbol{x} \\ &= \langle \mathcal{R}^* g,\, f\rangle, \end{aligned}$$

meaning

$$\mathcal{R}^* g(\boldsymbol{x}) = \int_0^\pi g(\langle \boldsymbol{x},\, \boldsymbol{n}_\theta\rangle, \theta)d\theta.$$

Replacing g by $\mathcal{R}g$, we recover the backprojection formula (modulo a factor of π):

$$\mathcal{R}^*\mathcal{R}g(\boldsymbol{x}) = \int_0^\pi \mathcal{R}g(\langle \boldsymbol{x},\, \boldsymbol{n}_\theta\rangle, \theta)d\theta = \pi\tilde{g}(\boldsymbol{x}).$$

Compare this with the following result.

Proposition 1. *We have*

$$\widehat{\mathcal{R}^*\mathcal{R}g}(\boldsymbol{\omega}) = \frac{2\pi}{\|\boldsymbol{\omega}\|}\hat{g}(\boldsymbol{\omega}). \tag{3}$$

We see from this proposition that

$$\widehat{\tilde{g}}(\omega) = \frac{2}{\|\boldsymbol{\omega}\|}\hat{g}(\boldsymbol{\omega}),$$

which shows $\tilde{g}$ (when properly rescaled) is obtained from g by attenuating high frequencies, namely by weights $\|\boldsymbol{\omega}\|^{-1}$. This matches our observation that $\tilde{g}$ is a smoothed version of g, since faster decay in the frequency domain translates to greater regularity in the spatial domain.

Proof. Let f, g be smooth and rapidly decaying. Recall from Lecture 9 the projection-slice theorem:

$$\int \mathcal{R}f(t,\theta)e^{-irt}\, dt = \hat{f}(r\boldsymbol{n}_\theta). \tag{4}$$

To establish (3), we compute $\langle f,\, \mathcal{R}^*\mathcal{R}g\rangle$ in two ways. On one hand, the Parseval-Plancherel theorem yields

$$\langle f,\, \mathcal{R}^*\mathcal{R}g\rangle = \frac{1}{(2\pi)^2}\langle \hat{f},\, \widehat{\mathcal{R}^*\mathcal{R}g}\rangle = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^\infty \overline{\hat{f}(r\boldsymbol{n}_\theta)}\widehat{\mathcal{R}^*\mathcal{R}g}(r\boldsymbol{n}_\theta) r\, drd\theta.$$

On the other hand, it also gives

$$\begin{aligned}\langle f, \mathcal{R}^*\mathcal{R}g\rangle = \langle \mathcal{R}f, \mathcal{R}g\rangle &= \int_0^\pi \int_{\mathbb{R}} \overline{\mathcal{R}f(t,\theta)}\mathcal{R}g(t,\theta)\; dtd\theta \\ &= \int_0^\pi \langle \mathcal{R}f(\,\cdot\,,\theta),\, \mathcal{R}g(\,\cdot\,,\theta)\rangle\; d\theta \\ &= \frac{1}{2\pi}\int_0^\pi \langle \widehat{\mathcal{R}f}(\,\cdot\,,\theta),\, \widehat{\mathcal{R}g}(\,\cdot\,,\theta)\rangle\; d\theta \\ &= \frac{1}{2\pi}\int_0^\pi \int_{-\infty}^{\infty} \overline{\hat{f}(r\boldsymbol{n}_\theta)}\hat{g}(r\boldsymbol{n}_\theta)\; drd\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} \int_0^{\infty} \overline{\hat{f}(r\boldsymbol{n}_\theta)}\hat{g}(r\boldsymbol{n}_\theta)\; drd\theta,\end{aligned}$$

where we have used (4). Consequently,

$$0 = \int_0^\infty \int_0^{2\pi} \overline{\hat{f}(r\boldsymbol{n}_\theta)}\left[\widehat{\mathcal{R}^*\mathcal{R}g}(r\boldsymbol{n}_\theta)r - 2\pi\hat{g}(r\boldsymbol{n}_\theta)\right] drd\theta = \iint \overline{\hat{f}(\boldsymbol{\omega})}\left[\widehat{\mathcal{R}^*\mathcal{R}g}(\boldsymbol{\omega}) - \frac{2\pi}{\|\boldsymbol{\omega}\|}\hat{g}(\boldsymbol{\omega})\right] d\boldsymbol{\omega},$$

for all suitable f and g. This proves the claim. □

Consider the case when $g(\boldsymbol{x}) = e^{i\langle\boldsymbol{\omega}_0,\, \boldsymbol{x}\rangle}$. The proposition tells us

$$\widehat{\mathcal{R}^*\mathcal{R}g}(\boldsymbol{\omega}) = \frac{2\pi}{\|\boldsymbol{\omega}\|}\hat{g}(\boldsymbol{\omega}) = \frac{2\pi}{\|\boldsymbol{\omega}\|}\delta_{\boldsymbol{\omega}_0} = \frac{2\pi}{\|\boldsymbol{\omega}_0\|}\delta_{\boldsymbol{\omega}_0},$$

which gives

$$\mathcal{R}^*\mathcal{R}g = \frac{2\pi}{\|\boldsymbol{\omega}_0\|}e^{i\langle\boldsymbol{\omega}_0,\, \boldsymbol{x}\rangle}.$$

The complex exponential $e^{i\langle\boldsymbol{\omega}_0,\, \boldsymbol{x}\rangle}$ is thus an eigenfunction of $\mathcal{R}^*\mathcal{R}$ with eigenvalue $2\pi/\|\boldsymbol{\omega}_0\|$. From this, we conclude that the singular values of $\mathcal{R}$ are $\sqrt{2\pi/\|\boldsymbol{\omega}\|}$. Since these singular values tend to 0, $\mathcal{R}^{-1}$ is unbounded. We have thus provided a quantitative statement about the ill-possedness of inverting the Radon transform.

We can express this in the language of linear algebra. Consider the singular value decomposition of $\mathcal{R}$:

$$\mathcal{R} = U\Sigma V^*,$$

where U and V are unitary, and Σ is diagonal. In particular, Σ consists of the singular values of $\mathcal{R}$ (the square roots of eigenvalues of $\mathcal{R}^*\mathcal{R}$), and the columns of V are the corresponding eigenvectors. From the above, we know these eigenvectors are complex exponentials, and so $\hat{f} = V^*f$ is the Fourier coefficients of f. Supposing $\mathcal{R}f$ is observed with noise z as $g = \mathcal{R}f + z$, we have in the frequency domain

$$U^*g = \Sigma\hat{f} + U^*z,$$

where Σ is a multiplication operator. It is now apparent that the ill-posedness of Radon inversion comes from the ill-posedness of a deconvolution. Just as in (2), the inversion formula

$$\Sigma^{-1}U^*g = \hat{f} + \Sigma^{-1}U^*z$$

is plagued by the unbounded growth in the entries of Σ^{-1}.

5 Regularized inversion

We now understand the inversion problem as formally solved, but whose solution is infeasible in practice. So what are we to do, if the solution itself is not really a solution? Without any other knowledge of the signal f, we are truly hopeless! But in just about every application imaginable, we have some idea of properties that f should possess. To start, any physically realized signal should have finite energy. So it is reasonable to expect that the minimizer of

$$\min_f \; \frac{1}{2}\|g - \mathcal{R}f\|^2 + \lambda\|f\|^2 \tag{5}$$

is a good approximation to the true signal, for suitably chosen λ. That is, instead of performing direct inversion, we could solve the optimization problem (5) to yield an estimate f of the underlying signal, whose Radon transform $\mathcal{R}f$ is close to the observed data g and whose energy $\|f\|^2$ is not too large.

We can generalize this strategy by penalizing something other than energy. In general, we call the proposed optimization the **regularized inversion problem**:

$$\min_f \; \underbrace{\frac{1}{2}\|g - \mathcal{R}f\|^2}_{\text{goodness of fit}} + \underbrace{\lambda P(f)}_{\text{regularization}}$$

There is evidently a tradeoff between **data fidelity** and **complexity**, and the parameter λ determines their relative importance. There are many popular choices for the **penalty** P, each of which reflects a particular belief about the form of the true signal. There is no one-size-fits-all; the best choice for P is application-specific and must penalize (or reward, in a different perspective) a feature distinctive of the signals under consideration. Here are some examples:

- $P(f) = \|f\|^2$, which penalizes energy as above. This is called **Tikhonov regularization**. One can also propose a **constrained** optimization:

  $$\min_f \; \|f\| \quad \text{subject to} \quad \|g - \mathcal{R}f\| \le \delta,$$

  where δ is a tolerance on the disparity between observed and predicted data. If there are many possible choices of f with $\mathcal{R}f$ within this tolerance, then this problem chooses a solution with minimum L^2-norm.

- $P(f) = \|\nabla f\|_2$, which penalizes the energy of the derivative[2] of f. Since the derivative is not allowed to be too large, this penalty function enforces additional smoothness in the reconstructed signal.

- $P(f) = \|f\|_{\mathrm{TV}}$, the **total variation** of f, given by

  $$\|f\|_{\mathrm{TV}} = \|\nabla f\|_1 = \int |\nabla f(\boldsymbol{x})|\, d\boldsymbol{x}.$$

  While L^2-regularization such as $P(f) = \|\nabla f\|_2$ seems a natural choice for the recovery of images, this L^1-regularization tends perform better. Intuitively, penalizing the L^1-norm of

[2] Even when f is not differentiable in the classical sense, norms of its derivatives can be understood in the sense of distributions. See Lecture 3 for a discussion of distributional derivatives.

the derivative removes noise artifacts by enforcing continuity, but still permits the sharp edge discontinuities usually present in physical images. See [7, 1, 3] for some of the influential work in this direction. The heuristic observation is that except for the contrasts created by different objects in the image, adjacent pixels are generally of the same tone. That is, images of interest (e.g. medical scans) tend to resemble piecewise continuous functions, which are preferred by TV-regularization but not by L^2-regularization.

- $P(f) = \|Wf\|_1$, where W is a transformation under which f should be **sparse**. For instance, perhaps we know *a priori* that the signal should have a sparse representation with respect to a particular orthonormal basis or spanning set (e.g. Fourier basis, wavelet basis, curvelet basis), and then W would represent f in that basis (notice that the L^2-norm would not suffice in this case, as the change-of-basis transformation would preserve $\|f\|_2$). This type of regularization is commonly called **basis pursuit** (see [2]).
- One can use a combination of several penalties designed for a specific application. Study of these approaches is usually spurned by experimental success.

For the particular case of $P(f) = \|f\|^2$, one can check via calculus and convexity that the solution to the regularized inversion problem is

$$f_{\text{rec}} = \left(\mathcal{R}^*\mathcal{R} + \lambda I\right)^{-1} \mathcal{R}^* g.$$

Using Proposition 1, we see that

$$\hat{f}_{\text{rec}}(\boldsymbol{\omega}) = \left(\frac{2\pi}{\|\boldsymbol{\omega}\|} + \lambda\right)^{-1} \widehat{\mathcal{R}^* g}(\boldsymbol{\omega}) = \frac{1}{2\pi} \cdot \frac{\|\boldsymbol{\omega}\|}{1 + \frac{\lambda}{2\pi}\|\boldsymbol{\omega}\|} \widehat{\mathcal{R}^* g}(\boldsymbol{\omega}).$$

Here the weight that appears in the reconstruction, namely

$$\frac{\|\boldsymbol{\omega}\|}{1 + \frac{\lambda}{2\pi}\|\boldsymbol{\omega}\|},$$

is *bounded*. This mitigates the problem of amplifying high frequencies by an arbitrarily large factor.

No matter the regularization chosen, consideration must be given to the computational tractability of solving the minimization problem. No matter how accurately a penalty function describes the desired reconstruction, it is not useful unless it gives an optimization problem that can be numerically solved quickly. This aspect of regularization is a field in itself. To see implementations of some of the regularizations we have discussed, the interested reader might start by browsing [4].

6 A comment on physical assumptions

We shall conclude our discussion of CT by revisiting the assumptions made in our model (refer to Lecture 8). In particular, we assumed all incident X-rays have the same frequency. This is physically inaccurate, of course, and so we should ask how our methods break down when this assumption is violated. At a given energy E, Beer's Law gives that

$$\log \frac{I_0(E)}{I_1(E)} = -\int_{\text{path}} \mu(x, E)\ dx, \tag{6}$$

where I_0 and I_1 are the incoming and outgoing flux, respectively, and μ is the attenuation coefficient of the material intersecting the X-ray trajectories. But this identity only applies to a *fixed* energy. For incident X-rays having variable energy (equivalently, variable frequency), the incident flux is given by some distribution $S(E)\,dE$. Then the true relationship becomes

$$\frac{I_0}{I_1} = \int_{\text{energies}} S(E)\,\exp\left(-\int_{\text{path}} \mu(x,E)\,dx\right)\,dE. \tag{7}$$

Before S was taken to be a delta function, in which case (7) reduces to (6). Now the problem is no longer linear in μ. Our approach of using the linear transformation $\mathcal{R}$ is thus problematic. This nonlinear problem is not necessarily impossible to solve, but the methods used would certainly need to be more sophisticated.

References

[1] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.

[2] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.

[3] David L Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[4] Tom Goldstein and Stanley Osher. The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[5] Alexander Hertle. On the problem of well-posedness for the radon transform. In *Mathematical Aspects of Computerized Tomography*, pages 36–44. Springer, 1981.

[6] SI Kabanikhin. Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-Posed Problems*, 16(4):317–357, 2008.

[7] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.