Inverse problems

Regularization
Lecture 3

Nikolai Piskunov

2014

Regularization

- An ill-posed problem can be made well-posed by restricting a *subset* of possible solutions

- Two questions:
  - How to make sure the subset contains only one solution?
  - Will that be the "correct" one?

- We can choose a strategy and try to answer those question for specific method.

Two methods:

1. **Search for quasi-solution**. Quasi-solution is the solution that is searched for on a subset of all possible solutions.

2. **Replacing the original problem with a similar well-posed problem**. We replace the kernel with the one that gives very close g's for the same f's but is well behaved.

- **Quasi-solutions.**
  One typical problem that is often solved with this method is deconvolution or the removal of the instrumental profile. The original signal is registered by an equipment described with a PSF $K(x)$:

$$g(x) = \int_{-\infty}^{+\infty} K(x-y) f(y) dy$$

  The PSF (kernel) usually quickly drops away from the center but never goes below zero.

In many cases a <u>Gaussian</u> is a good model for the PSF:

$$K(x) = \frac{1}{\sqrt{\pi}\delta}\, e^{-(x/\delta)^2}$$

The Fourier transform is defined as:

$$\bar{f}(\omega) = \mathrm{F}\, f \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t)e^{i\omega t}\, dt$$

$$f(t) = \mathrm{F}^{-1}\, f \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \bar{f}(\omega)e^{-i\omega t}\, d\omega$$

One interesting property of Fourier transform is relevant for the convolution:

$$\bar{u}(\omega) \cdot \bar{v}(\omega) = F\left[\int_{-\infty}^{+\infty} u(x-y)v(y)dy\right]$$

Another interesting property is that the Fourier transform of a Gaussian is a Gaussian. Assuming that we know the instrumental profile (PSF) we can formally perform deconvolution:
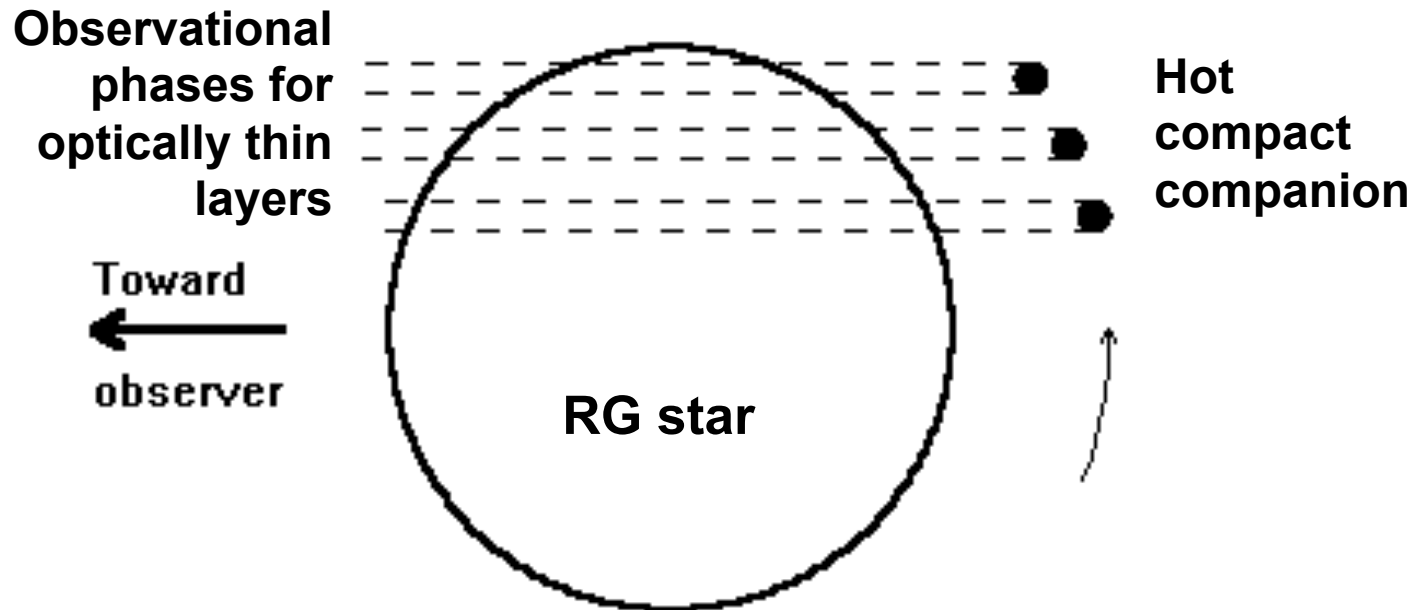
$$f(x) = F^{-1}\left[Fg/FK\right]$$

In practice, the $\bar{g}$ and $\bar{K}$ are known on a discrete grid with certain measurement errors. The noise has frequency-independent spectrum, so the Fourier transform of the kernel and the measurements would look something like this:

Ratio $\bar{g}/\bar{K}$ →

Measurements $\bar{g}$
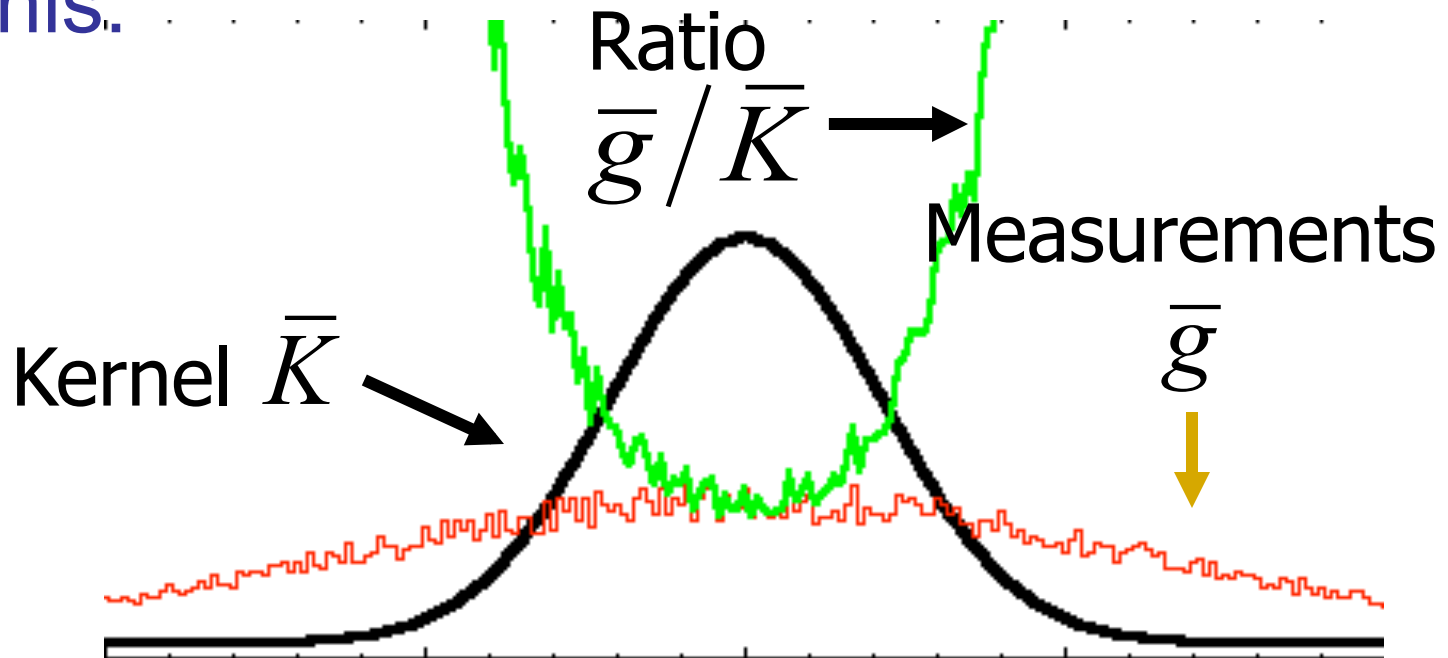
Kernel $\bar{K}$

Quasi-solution: Let's restrict ourselves with solutions containing only low frequencies, for example, from 0 up to the point where the transform of the kernel becomes equal to $\overline{g}$. In more sophisticated cases one may try to create a model for the noise spectrum and use that for the cut-off. The solution is stable and unique and it will actually approach the true f, as the S/N is improved.

Note, that narrow spectrum of **K** means that the kernel decays slowly making our constraints on the solution more strict.

One excellent example of use of Quasi-solution method: the study of extended atmosphere of red giant stars in eclipsing binary systems. We observe light curves of the secondary as it shines through the atmosphere of the giant star and derive T and P as functions of radius. Unique solution is found if T and P are assumed to be monotonous functions of the radius decreasing outwards.

Observational phases for optically thin layers

Toward

observer

RG star

Hot compact companion

Let look at the iterations:

- Get T(r) and P(r) from a model atmosphere

- Compute molecular-ionization equilibrium to get partial number densities

- Get opacities and optical path

- Adjust source function, get new temperature structure

- Re-compute model atmosphere

- Iterate

Regularized functional

- We are going to replace the functional Φ which results in an ill-posed problem with another functional which is close, has better properties and behaves in a predictable way.

- Strictly speaking we want to replace $\Phi(f)$ with $\Omega(f, \Lambda)$ where for measurement error $\delta > 0$ there are numbers Λ and $\varepsilon > 0$ such that for all measurements within the error bars $\left\| g - g^{true} \right\| < \delta$ the solutions minimizing Ω form a compact set around the true solution:

$$\left\| f - f^{true} \right\| < \varepsilon$$

One example of using regularization in numerical methods is the approximate calculation of derivative:

$$f(x) = \frac{dg}{dx} \approx \frac{g(x + \Delta) - g(x)}{\Delta}$$

Assuming that a sequence of f will converge to the true derivative as $\Delta \rightarrow 0$ we find an ill-posed problem. If a numerical representation can be split in a true value and rounding error $g^m = g^t + e$ then the approximation can be re-written as:

$$\frac{g^m(x + \Delta) - g^m(x)}{\Delta} = \frac{g^t(x + \Delta) - g^t(x)}{\Delta} + \frac{e(x)}{\Delta}$$

and $\dfrac{g^t(x + \Delta) - g^t(x)}{\Delta} \xrightarrow{\Delta \rightarrow 0} \dfrac{dg}{dx}$ but $\dfrac{\delta}{\Delta} \rightarrow \infty$

- In order to keep our estimate reasonable we have to keep $\Delta = \Lambda(\delta) > \delta$.
  For smaller Δ the errors will dominate our estimates.

- We can generalize this idea and introduce a positive functional $\mathbf{R}(f, \Lambda)$ that is defined for all f's and has a property that for a given d only a compact subset of f's fulfills the condition: $\mathbf{R}(f, \Lambda) < d$

- The compactness means that for a given d any pair of f's such that: $\mathbf{R}(f_1) < d$ and $\mathbf{R}(f_2) < d$ is closer than certain $\varepsilon > \|f_1 - f_2\|$ and for $d \to 0,\ \varepsilon \to 0$

- In 1943 Tikhonov proved that if such $\mathbf{R}$ can be found

$$\Omega(f, \Lambda) \equiv \Phi(f) + \Lambda \cdot \mathbf{R}(f)$$

  is a regularized functional, stable against small perturbations in g's.

- The idea behind is that we may have many solutions f such that $\left\| \mathbf{K}f - g^m \right\| < \delta$ but only a compact number of them will result in small $\mathbf{R}$'s. If more then one solution exists they would be close to each other and Λ can be selected such as to make the solution unique.

R is called the <u>regularizing functional</u>. Now instead of solving the problem searching for minimum of Φ, we will search for the minimum of Ω. One can also think of this as of conditional minimization where the relative weight of additional condition(s) is determined by Lagrangian multiplier Λ also known as the <u>regularization parameter</u>.

We would like to find the regularization function such that it will make the search for minimum of Ω a well-posed problem while Λ can be adjusted depending on measurements to ensure uniqueness.

3 big questions:

- Is there a general form of $\mathbf{R}$ that will ensure the stability and uniqueness of the solution?

- How do we choose the regularization parameter Λ?

- Can we say anything about the relation between the properties of g's and resulting solution for f's minimizing Ω?

Tikhonov theorem: regularizing functional

In 1D case for functions defined on a finite interval:

$$\mathbf{K}f = \int_a^b K(x,y)f(y)dy = g(x)$$

the functional:

$$\mathbf{R}(f) = \int_a^b \left[p(y)f^2 + q(y)\left(\frac{df}{dy}\right)^2 \right] dy$$

with any non-negative functions $q(x)$ and $p(x)$ regularizes the problem.

Can we prove that? For a linear operator: **Yes**. Suppose we have two *different* solutions f_1 and f_2 that minimize Ω. Let's construct a linear combination: $f = f_1 + \beta \cdot (f_2 - f_1)$

The regularized functional becomes now a simple quadratic function of β:

$$\Omega(\beta) = \left\| \mathrm{K}(f) - g \right\| + \mathrm{R}(f) = A + B\beta + C\beta^2$$

with:

$$C = \int_a^b \left\{ \int_a^b K(x,y) \cdot (f_2 - f_1) \, dy \right\}^2 dx +$$

$$+ \int_a^b \left[p \cdot (f_2 - f_1)^2 + q \cdot (f_2' - f_1')^2 \right] dy > 0$$

Thus $\Omega(\beta)$ has only one minimum!

- In case of nonlinear kernel the prove is not that easy. One can define special classes of kernels for which the uniqueness can be proven. One class important for astrophysical applications is when the kernel is a monotonous in f.

- The regularization parameter Λ in theory can be predicted if the error bars of the measurements are known. Unfortunately, in practice the model often contains additional tuning parameters, systematic errors etc. Instead we adjust Λ as we approach the solution.

- The regularizing functional above is called Tikhonov regularization. Other regularizing functional forms:
  - Maximum entropy (MEM)
  - Maximum likelyhood

- MEM. Entropy in information theory has been introduced by Shannon in 1948 as the information content of the probability distribution: $\mathbf{S} = -\sum p_i \log p_i / m_i$
  The information content does not depend on the physical units so p_i are unitless, e.g.:

$$p_i = \frac{f_i}{\frac{1}{N} \sum f_i}$$

- The regularization is achieved by <u>maximizing the entropy</u> so the regularized functional will be:

$$\Omega(f) = \|Kf - g\| + \Lambda \sum p_i \log p_i$$

- **Home work:**
  Implement Tikhonov regularization assuming $p(x) = 0$ and $q(x) = 1$ and MEM regularization.