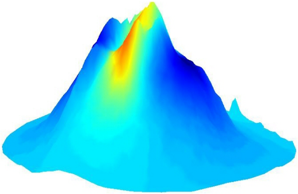
Motivation: Why Inverse Problems?

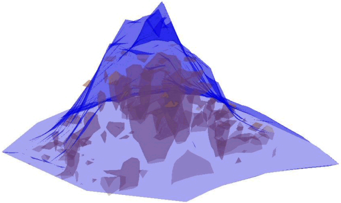
A large-scale example, coming from a collaboration with Università degli Studi di Napoli "Federico II" in Naples.

From measurements of the magnetic field above Vesuvius, we determine the activity inside the volcano.



Measurements
on the surface

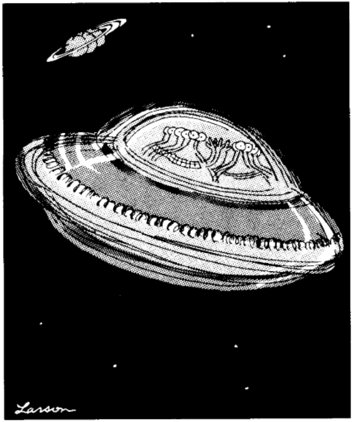
$\Rightarrow$



Reconstruction
inside the volcano

Another Example: the Hubble Space Telescope

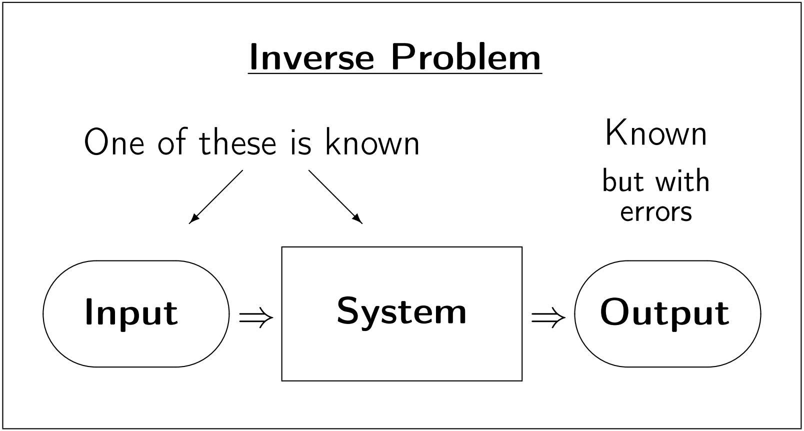
For several years, the HST produced blurred images.



Another photograph from the Hubble telescope

Inverse Problems

. . . typically arise when one wants to compute information about some "interior" properties using "exterior" measurements.



Inverse Problems: Examples

A quite generic formulation:

$$\boxed{\int_\Omega \text{input} \times \text{system} \; d\Omega = \text{output}}$$

Image restoration

$$\text{scenery} \to \text{lens} \to \text{image}$$

Tomography

$$\text{X-ray source} \to \text{object} \to \text{damping}$$

Seismology

$$\text{seismic wave} \to \text{layers} \to \text{reflections}$$

Discrete Ill-Posed Problems

Our generic ill-posed problem:

A Fredholm integral equation of the first kind

$$\int_0^1 K(s,t)\, f(t)\, dt = g(s) \, , \qquad 0 \le s \le 1 \; .$$

Definition of a discrete ill-posed problem (DIP):

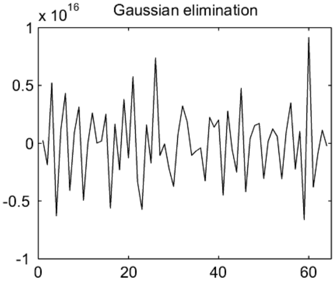
1. a square or over-determined system of linear algebraic equations

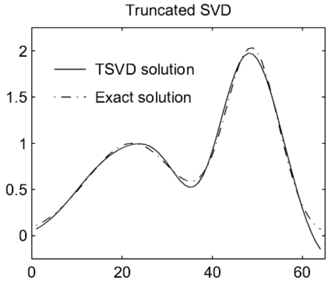
$$A\,x = b \qquad \text{or} \qquad \min_x \|A\,x - b\|_2$$

2. whose coefficient matrix A has a huge condition number, and
3. comes from the discretization of an inverse/ill-posed problem.

Computational Issues

The plots below show solutions x to the 64×64 DIP $A\,x = b$.





- Standard numerical methods (`x = A\b`) produce useless results.
- Specialized methods (this course) produce "reasonable" results.

The Mechanisms of Ill-Conditioned Problems

Consider a linear system with coefficient matrix and right-hand side

$$A = \begin{pmatrix} 0.16 & 0.10 \\ 0.17 & 0.11 \\ 2.02 & 1.29 \end{pmatrix}, \qquad b = \begin{pmatrix} 0.27 \\ 0.25 \\ 3.33 \end{pmatrix} = A \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0.01 \\ -0.03 \\ 0.02 \end{pmatrix}.$$

There is no vector x such that $A\,x = b$.

The least squares solution, which solves the problem

$$\min_x \|A\,x - b\|_2,$$

is given by

$$x_{\mathrm{LS}} = \begin{pmatrix} 7.01 \\ -8.40 \end{pmatrix} \qquad \Rightarrow \qquad \|A\,x_{\mathrm{LS}} - b\|_2 = 0.022\ .$$

Far from exact solution $(\,1\,,\,1\,)^T$ yet the residual is small.

Other Solutions with Small Residual

Two other "solutions" with a small residual are

$$x' = \begin{pmatrix} 1.65 \\ 0 \end{pmatrix} \qquad \Rightarrow \qquad \|A\,x' - b\|_2 = 0.031$$

$$x'' = \begin{pmatrix} 0 \\ 2.58 \end{pmatrix} \qquad \Rightarrow \qquad \|A\,x'' - b\|_2 = 0.036 \ .$$

All the "solutions" x_{LS}, x' and x'' have small residuals, yet they are far from the exact solution!

What we have learned:

- The matrix A is ill conditioned.
- Small perturbations of the data (here: b) can lead to large perturbations of the solution.
- A small residual does not imply a good solution.

(All this is well known stuff from matrix computations.)

Stabilization!

It turns out that we can modify the problem such that the solution is more stable, i.e., less sensitive to perturbations.

Example: enforce an upper bound on the solution norm $\|x\|_2$:

$$\min_x \|A\,x - b\|_2 \qquad \text{subject to} \qquad \|x\|_2 \leq \delta \ .$$

The solution x_δ depends in a nonlinear way on δ:

$$x_{0.1} = \begin{pmatrix} 0.08 \\ 0.05 \end{pmatrix} \ , \qquad x_1 = \begin{pmatrix} 0.84 \\ 0.54 \end{pmatrix}$$

$$x_{1.385} = \begin{pmatrix} 1.17 \\ 0.74 \end{pmatrix} \ , \qquad x_{10} = \begin{pmatrix} 6.51 \\ -7.60 \end{pmatrix} \ .$$

By supplying valuable additional information we can compute a good approximate solution.

Inverse Problems $\rightarrow$ Ill-Conditioned Problems

Whenever we solve an inverse problem on a computer, we face difficulties because the computational problems are ill conditioned.

The purpose of these lectures are:

1. To explain why ill-conditioned computations always arise when solving inverse problems.
2. To explain the fundamental "mechanisms" underlying the ill conditioning.
3. To explain how we can modify the problem in order to stabilize the solution.
4. To show how this can be done efficiently on a computer.

Regularization methods is at the heart of all this.

Inverse Problems are Ill-Posed Problems

Hadamard's definition of a *well-posed problem* (early 20th century):

1. the problem must have a solution,
2. the solution must be unique, and
3. it must depend continuously on data and parameters.

If the problem violates any of these requirements, it is *ill posed*.

Condition 1 can be fixed by reformulating/redefining the solution.

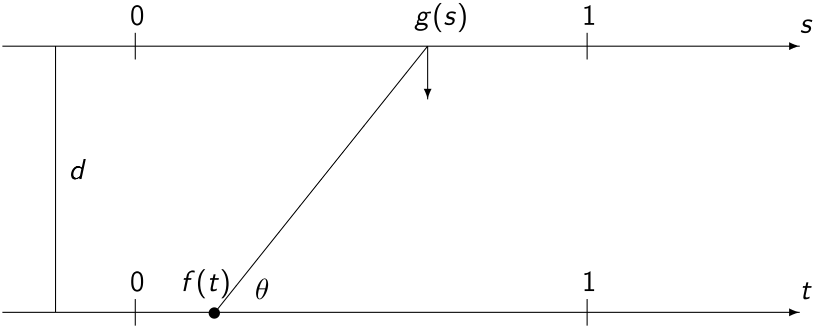
Condition 2 can be "fixed" by additional requirements to the solution, e.g., that of minimum norm.

Condition 3 is harder to "fix" because it implies that

- arbitrarily small perturbations of data and parameters can produce arbitrarily large perturbations of the solution.

Model Problem: Gravity Surveying

- Unknown mass density distribution $f(t)$ at depth d below surface, from 0 to 1 on t axis.
- Measurements of vertical component of gravitational field $g(s)$ at surface, from 0 to 1 on the s axis.



Setting Up the Integral Equation

The value of $g(s)$ due to the part dt on the t axis

$$dg = \frac{\sin\theta}{r^2}\, f(t)\, dt \ ,$$

where $r = \sqrt{d^2 + (s-t)^2}$. Using that $\sin\theta = d/r$, we get

$$\frac{\sin\theta}{r^2}\, f(t)\, dt = \frac{d}{(d^2 + (s-t)^2)^{3/2}}\, f(t)\, dt \ .$$

The total value of $g(s)$ for $0 \le s \le 1$ is therefore

$$g(s) = \int_0^1 \frac{d}{(d^2 + (s-t)^2)^{3/2}}\, f(t)\, dt \ .$$

This is the $\boxed{\text{forward problem.}}$

Our Integral Equation

Fredholm integral equation of the first kind:

$$\int_0^1 \frac{d}{(d^2 + (s-t)^2)^{3/2}} \, f(t) \, dt = g(s) \ , \qquad 0 \le s \le 1 \ .$$

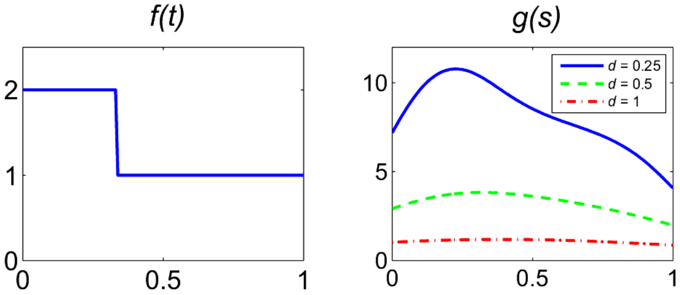
The kernel K, which represents the model, is

$$K(s,t) = h(s-t) = \frac{d}{(d^2 + (s-t)^2)^{3/2}} \ ,$$

and the right-hand side g is what we are able to measure.

From K and g we want to compute f, i.e., an $\boxed{\text{inverse problem.}}$

Numerical Examples



Observations:

- The signal/"data" $g(s)$ is a smoothed version of the source $f(t)$.
- The deeper the source, the weaker the signal.
- The discontinuity in $f(t)$ is not visible in $g(s)$.

Fredholm Integral Equations of the First Kind

Our generic inverse problem:

$$\int_0^1 K(s,t)\, f(t)\, dt = g(s), \qquad 0 \le s \le 1 \ .$$

Here, the *kernel* $K(s,t)$ and the *right-hand side* $g(s)$ are known functions, while $f(t)$ is the unknown function.

In multiple dimensions, this equation takes the form

$$\int_{\Omega_{\mathbf{t}}} K(\mathbf{s},\mathbf{t})\, f(\mathbf{t})\, d\mathbf{t} = g(\mathbf{s}), \qquad \mathbf{s} \in \Omega_{\mathbf{s}} \ .$$

An important special case: $K(s,t) = h(s-t) \rightarrow$ *deconvolution:*

$$\int_0^1 h(s-t)\, f(t)\, dt = g(s), \qquad 0 \le s \le 1$$

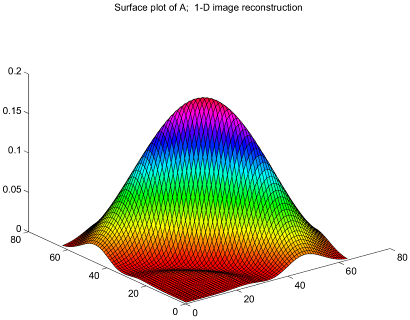
(and similarly in more dimensions).

Another Example: 1-D Image Restoration

Kernel K: point spread function for an infinitely long slit of width one wavelength.

Independent variables t and s are the angles of the incoming and scattered light.

Regularization Tools: `shaw`.



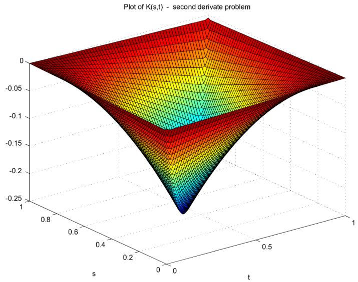
$$
\begin{aligned}
K(s,t) &= \big(\cos(s)+\cos(t)\big)^2 \left(\frac{\sin(u)}{u}\right)^2 \\
u &= \pi\,\big(\sin(s)+\sin(t)\big)
\end{aligned}
$$

$$
\int_{-\pi/2}^{\pi/2} K(s,t)\, f(t)\, dt = g(s)\ , \quad -\pi/2 \le s \le \pi/2
$$

Yet Another Example: Second Derivative

Kernel K: Green's function for the second derivative

$$K(s,t) = \begin{cases} s(t-1)\ , & s < t \\ t(s-1)\ , & s \geq t \end{cases}$$



Regularization Tools: `deriv2`.

Not differentiable across the line $t = s$.

$$\int_0^1 K(s,t)\, f(t)\, dt = g(s)\ , \quad 0 \leq s \leq 1$$

Solution:

$$f(t) = g''(t)\ , \qquad 0 \leq t \leq 1\ .$$

The Riemann-Lebesgue Lemma

Consider the function

$$f_p(t) = \sin(2\pi p\, t) \ , \qquad p = 1, 2, \ldots$$

then for $p \to \infty$ and "arbitrary" K we have

$$g_p(s) = \int_0^1 K(s,t)\, f_p(t)\, dt \to 0 \ .$$

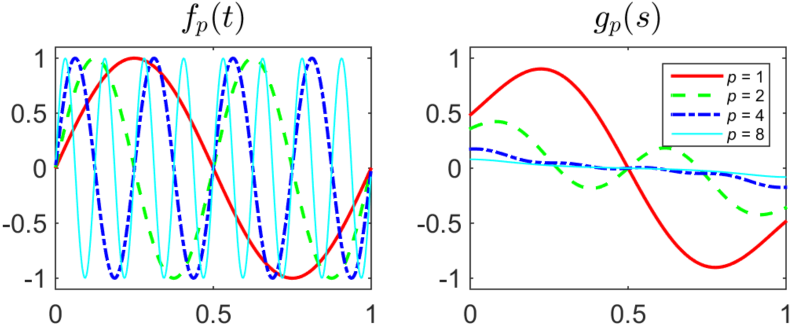
Smoothing: high frequencies are damped in the mapping $f \mapsto g$.

Hence, the mapping from g to f must amplify the high frequencies.

Therefore we can expect difficulties when trying to reconstruct f from noisy data g.

Illustration of the Riemann-Lebesgue Lemma

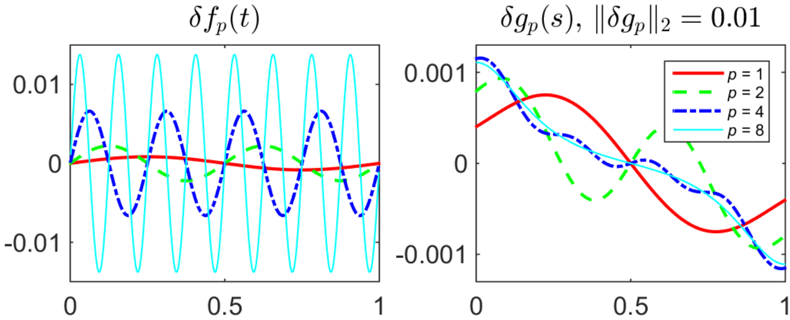
Gravity problem with $f_p(t) = \sin(2\pi p\, t)$, $p = 1, 2, 4,$ and 8.



Higher frequencies are damped more than low frequencies.

Difficulties with High Frequencies

In this example $\delta g_p(s) = \int_0^1 K(s,t)\,\delta f_p(t)\,dt$ and $\|\delta g_p\|_2 = 0.01$.



Higher frequencies are amplified more in the reconstruction process.

Why do We Care?

Why bother about these (strange) issues?

- Ill-posed problems model a variety of real applications:
  - Medical imaging (brain scanning, etc.)
  - Geophysical prospecting (search for oil, land-mines, etc.)
  - Image deblurring (astronomy, CSI[1], etc.)
  - Deconvolution of instrument's response.
- We can only hope to compute useful solutions to these problems if we fully understand their *inherent* difficulties . . .
- and how these difficulties carry over to the discretized problems involved in a computer solution,
- and how to deal with them in a satisfactory way.

[1] Crime Scene Investigation.

Some Important Questions

- How to discretize the inverse problem; here, the integral equation?
- Why is the matrix in the discretized problem always so ill conditioned?
- Why can we still compute an approximate solution?
- How can we compute it stably and efficiently?
- Is additional information available?
- How can we incorporate it in the solution scheme?
- How should we implement the numerical scheme?
- How do we solve large-scale problems?

The Singular Value Expansion (SVE)

For any square integrable kernel K holds

$$K(s,t) = \sum_{i=1}^{\infty} \mu_i \, u_i(s) \, v_i(t),$$

where $\langle u_i, u_j \rangle = \langle v_i, v_j \rangle = \delta_{ij}$, and $\mu_1 \geq \mu_2 \geq \mu_3 \geq \ldots \geq 0$.

The "fundamental relation" and the expansions

$$\int_0^1 K(s,t) \, v_i(t) \, dt = \mu_i \, u_i(s) \ , \qquad i = 1, 2, \ldots$$

$$f(t) = \sum_{i=1}^{\infty} \langle v_i, f \rangle \, v_i(t) \qquad \text{and} \qquad g(s) = \sum_{i=1}^{\infty} \langle u_i, g \rangle \, u_i(s)$$

lead to the expression for the solution:

$$f(t) = \sum_{i=1}^{\infty} \frac{\langle u_i, g \rangle}{\mu_i} \, v_i(t) \ .$$

The Singular Values

Ordering

$$\mu_1 \geq \mu_2 \geq \mu_3 \geq \cdots \geq 0 \ .$$

Norm of kernel

$$\|K\|^2 \equiv \int_0^1 \int_0^1 |K(s,t)|^2 \, ds \, dt = \sum_{i=1}^{\infty} \mu_i^2 \ .$$

Hence,

$$\mu_i = \mathcal{O}(i^{-q}) \ , \qquad q > 1/2$$

i.e., the μ_i decay faster than $i^{-1/2}$. We have:

- If the derivatives of order $0, \ldots, q$ exist and are continuous, then μ_i is approximately $\mathcal{O}(i^{-q-1/2})$.
- Second derivative: $\mu_i \approx i^{-2}$ ("moderately ill posed").
- 1-D image reconstruction: $\mu_i \approx e^{-2i}$ ("severely ill posed").

Example of SVE (Degenerate)

We can occasionally calculate the SVE analytically. Example

$$\int_{-1}^{1} (s+2t)\, f(t)\, dt = g(s), \qquad -1 \leq s \leq 1.$$

For this kernel $K(s,t) = s + 2t$ we have

$$\mu_1 = \mu_2 = 2/\sqrt{3}, \quad \mu_3 = \mu_4 = \ldots = 0.$$

$$u_1(s) = 1/\sqrt{2}, \quad u_2(s) = \sqrt{3/2}\, s$$

$$v_1(t) = \sqrt{3/2}\, t, \quad v_2(t) = 1/\sqrt{2}.$$

A solution exists only if

$$g \in \mathrm{range}(K) = \mathrm{span}\{u_1, u_2\},$$

i.e., if g is of the form

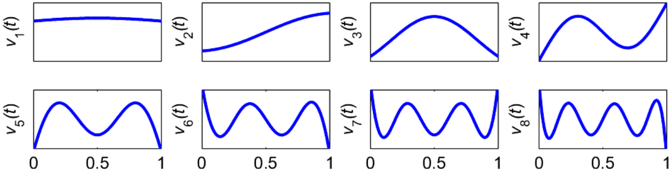
$$g(s) = c_1 + c_2\, s.$$

The Smoothing Effect

The "smoother" the kernel K, the faster the μ_i decay to zero:

- If the derivatives of order $0, \ldots, q$ exist and are continuous, then μ_i is approximately $\mathcal{O}(i^{-q-1/2})$.

The smaller the μ_i, the more oscillations (or zero-crossings) in the singular functions u_i and v_i.



Since $v_i(t) \to \mu_i \, u_i(s)$, higher frequencies are damped more than lower frequencies (smoothing) in the forward problem.

The Picard Condition

In order that there exists a square integrable solution f to the integral equation, the right-hand side g must satisfy

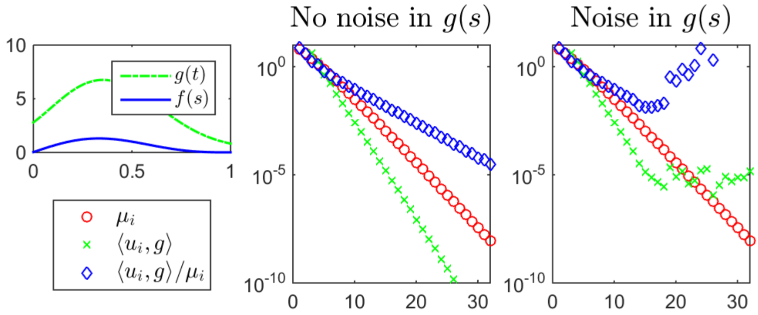
$$\sum_{i=1}^{\infty} \left(\frac{\langle u_i, g \rangle}{\mu_i} \right)^2 < \infty \ .$$

Equivalent condition: $g \in \text{range}(K)$.

In plain words: the absolute value of the coefficients (u_i, g) must decay *faster* than the singular values μ_i !

Main difficulty: a noisy g does not satisfy the PC!

Illustration of the Picard Condition



The violation of the Picard condition is the (simple) explanation of the instability of linear inverse problems in the form of first-kind Fredholm integral equations.

SVE analysis + Picard plot $\rightarrow$ insight $\rightarrow$ remedy $\rightarrow$ algorithms.

A Problem with no Solution

Ursell (1974) presented the following innocently-looking problem:

$$\int_0^1 \frac{1}{s+t+1}\, f(t)\, dt = 1, \qquad 0 \le s \le 1.$$

This problem *has no square integrable solution!*
Expand right-hand side $g(s) = 1$ in terms of the singular functions:

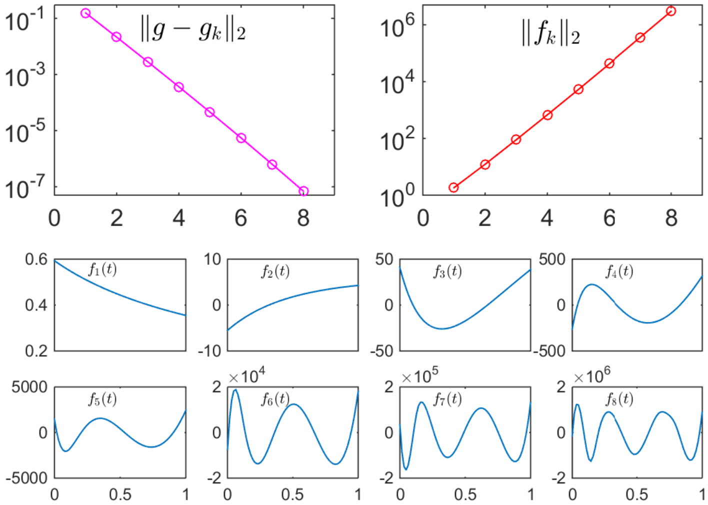
$$g_k(s) = \sum_{i=1}^{k} \langle u_i, g \rangle\, u_i(s); \qquad \|g - g_k\|_2 \to 0 \text{ for } k \to \infty.$$

Now consider $\int_0^1 \frac{f_k(t)}{s+t+1}\, dt = g_k(s)$, whose solution f_k is

$$f_k(t) = \sum_{i=1}^{k} \frac{\langle u_i, g \rangle}{\sigma_i}\, v_i(t).$$

Clearly $\|f_k\|_2$ is finite for all k; but $\|f_k\|_2 \to \infty$ for $k \to \infty$.

Ursell Problem – Numerical Results



Analytic SVEs are Rare

A few cases where analytic SVEs are available, e.g., the Radon transform.

But in most applications we must use numerical methods for analysis and solution of the integral equation.

Many of these lectures are devoted to numerical methods!

Our analysis has given us an understanding of the difficulties we are facing – and they will manifest themselves again in any numerical approach we're using to solve the integral equation.