Inverse Problems 3: Why Differentiation is Harder than Integration

Kurt Bryan

April 26, 2011

A More Interesting Problem

Put P_0 dollars at time $t = 0$ into a 401K with instantaneous return rate $r(t)$.

A More Interesting Problem

Put P_0 dollars at time $t = 0$ into a 401K with instantaneous return rate $r(t)$.

Forward Problem: Compute $P(t)$ from P_0 and $r(t)$. This means solving the DE

$$P'(t) = r(t)P(t)$$

A More Interesting Problem

Put P_0 dollars at time $t = 0$ into a 401K with instantaneous return rate $r(t)$.

Forward Problem: Compute $P(t)$ from P_0 and $r(t)$. This means solving the DE

$$P'(t) = r(t)P(t)$$

The solution is

$$P(t) = P_0 \exp\left(\int_0^t r(s)\,ds\right).$$

A More Interesting Problem

Inverse Problem: Estimate $r(t)$ from $P(t)$. This means finding $r(t)$ from the DE

$$P'(t) = r(t)P(t).$$

A More Interesting Problem

Inverse Problem: Estimate $r(t)$ from $P(t)$. This means finding $r(t)$ from the DE

$$P'(t) = r(t)P(t).$$

The solution is just

$$r(t) = P'(t)/P(t).$$

But it's not as simple as it looks...

Estimating the Interest Rate

Suppose we know $P(t)$ at times $t_k = k\Delta t$, $k = 0, 1, 2, \ldots$, rounded to the nearest penny of course. We can estimate

$$P'(t_k) \approx \frac{P(t_{k+1}) - P(t_{k-1})}{2\Delta t}$$

Estimating the Interest Rate

Suppose we know $P(t)$ at times $t_k = k\Delta t$, $k = 0, 1, 2, \ldots$, rounded to the nearest penny of course. We can estimate

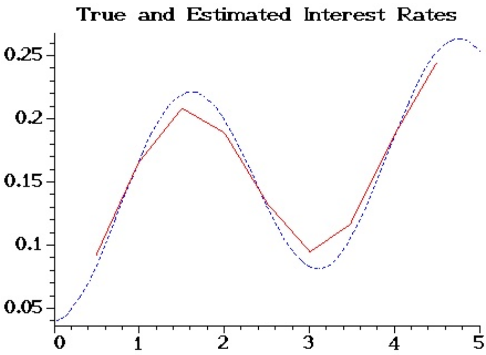
$$P'(t_k) \approx \frac{P(t_{k+1}) - P(t_{k-1})}{2\Delta t}$$

From $r(t) = P'(t)/P(t)$ we get

$$r(t_k) \approx \frac{P(t_{k+1}) - P(t_{k-1})}{2\Delta t P(t_k)}.$$

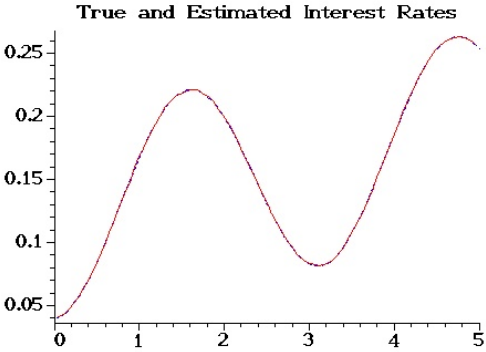
Example

Suppose $r(t) = 0.04(3 - 2\cos(2t) + t/3)$ on $0 \leq t \leq 5$, with $P(0) = 100$. If we use $r(t_k) \approx \frac{P(t_{k+1}) - P(t_{k-1})}{2\Delta t P(t_k)}$ with $\Delta t = 0.5$ the result is



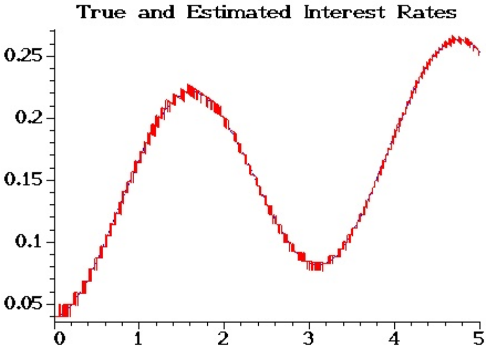
Example

With $\Delta t = 0.05$ the result is better:



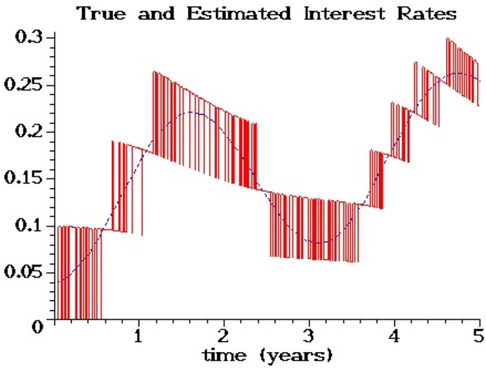
Example

But with $\Delta t = 0.005$ we get



Example

And $\Delta t = 0.0005$ yields



What's Wrong?

We don't really know $P(t_k)$, but $P(t_k)$ rounded to the nearest cent, i.e., we know

$$\tilde{P}(t_k) = P(t_k) + \epsilon_k$$

where $|\epsilon_k| \leq 0.005$ dollars.

What's Wrong?

We don't really know $P(t_k)$, but $P(t_k)$ rounded to the nearest cent, i.e., we know

$$\tilde{P}(t_k) = P(t_k) + \epsilon_k$$

where $|\epsilon_k| \leq 0.005$ dollars.
Our estimate of $P'(t_k)$ is then

$$P'(t_k) \approx \frac{\tilde{P}(t_{k+1}) - \tilde{P}(t_{k-1})}{2\Delta t}$$

What's Wrong?

We don't really know $P(t_k)$, but $P(t_k)$ rounded to the nearest cent, i.e., we know

$$\tilde{P}(t_k) = P(t_k) + \epsilon_k$$

where $|\epsilon_k| \leq 0.005$ dollars.
Our estimate of $P'(t_k)$ is then

$$\begin{aligned} P'(t_k) &\approx \frac{\tilde{P}(t_{k+1}) - \tilde{P}(t_{k-1})}{2\Delta t} \\ &= \underbrace{\frac{P(t_{k+1}) - P(t_{k-1})}{2\Delta t}}_{\text{better as } \Delta t \to 0} + \underbrace{\frac{\epsilon_{k+1} - \epsilon_{k-1}}{2\Delta t}}_{\text{may blow up as } \Delta t \to 0} . \end{aligned}$$

Computing f'

If we have noisy data $f_k = f(t_k) + \epsilon_k$ where ϵ_k is noise of fixed magnitude and approximate

$$f'(t_k) \approx \frac{f_{k+1} - f_k}{\Delta t}$$

Computing f'

If we have noisy data $f_k = f(t_k) + \epsilon_k$ where ϵ_k is noise of fixed magnitude and approximate

$$f'(t_k) \approx \frac{f_{k+1} - f_k}{\Delta t} = \underbrace{\frac{f(t_{k+1}) - f(t_k)}{\Delta t}}_{\text{better as } \Delta t \to 0} + \underbrace{\frac{\epsilon_{k+1} - \epsilon_k}{\Delta t}}_{\text{may blow up as } \Delta t \to 0}.$$

Computing f'

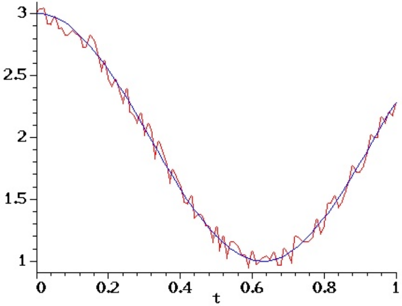
If we have noisy data $f_k = f(t_k) + \epsilon_k$ where ϵ_k is noise of fixed magnitude and approximate

$$f'(t_k) \approx \frac{f_{k+1} - f_k}{\Delta t} = \underbrace{\frac{f(t_{k+1}) - f(t_k)}{\Delta t}}_{\text{better as } \Delta t \to 0} + \underbrace{\frac{\epsilon_{k+1} - \epsilon_k}{\Delta t}}_{\text{may blow up as } \Delta t \to 0}.$$

Depending on the sign of $\epsilon_{k+1} - \epsilon_k$, the estimate of $f'(t_k)$ will be way high or low. The overall estimate of f' will be highly oscillatory.

Noisy f' Illustration

The blue curve is the graph of $f(t)$ (not $f'(t)$), the red a plot of the noisy sampled values f_k:



The formula $f'(t_k) \approx (f_{k+1} - f_k)/\Delta t$ estimates f' as the slope of the red curve.

Alternate Approach to Computing f'

Dumb Idea: Given data $f_0, f_1, \ldots, f_n$ (spacing Δt) estimate $f'(t_k)$ as d_k where the d_k minimize

$$Q(d_0, \ldots, d_{n-1}) = \sum_{k=0}^{n-1} \left(d_k - \frac{f_{k+1} - f_k}{\Delta t} \right)^2 .$$

Alternate Approach to Computing f'

Dumb Idea: Given data $f_0, f_1, \ldots, f_n$ (spacing Δt) estimate $f'(t_k)$ as d_k where the d_k minimize

$$Q(d_0, \ldots, d_{n-1}) = \sum_{k=0}^{n-1} \left(d_k - \frac{f_{k+1} - f_k}{\Delta t} \right)^2 .$$

But Q is a sum of squares—the minimum occurs when $d_k = (f_{k+1} - f_k)/\Delta t$ as before. This gives the exactly the same (bad) estimate as before.

An Improvement

Good Idea: Add to Q a term that (from the minimization perspective) penalizes highly oscillatory values for the d_k:

$$Q(d_0, \ldots, d_{n-1}) = \sum_{k=0}^{n-1} \left(d_k - \frac{f_{k+1} - f_k}{\Delta t} \right)^2$$

An Improvement

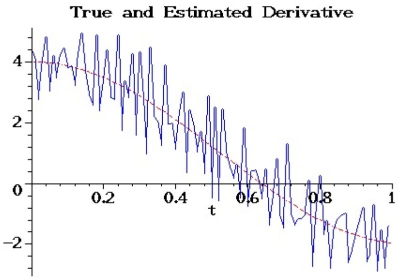
Good Idea: Add to Q a term that (from the minimization perspective) penalizes highly oscillatory values for the d_k:

$$Q(d_0, \ldots, d_{n-1}) = \sum_{k=0}^{n-1} \left(d_k - \frac{f_{k+1} - f_k}{\Delta t} \right)^2 + \underbrace{\frac{\alpha}{2} \sum_{k=0}^{n-2} \left(\frac{d_{k+1} - d_k}{\Delta t} \right)^2}_{\text{regularization term}}.$$

The parameter α is called the *regularization parameter*. We can adjust it.

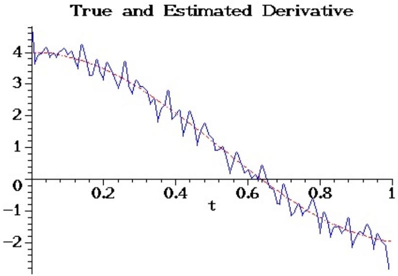
Example

Actual function $f(t) = t + \sin(3t)$ on $0 \leq t \leq 1$ sampled at 100 points, noise of magnitude 0.01 added to each sample. Here's the straight ($\alpha = 0$) estimate of f'.



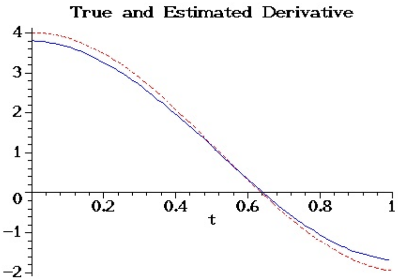
Example

The result with $\alpha = 10^{-4}$



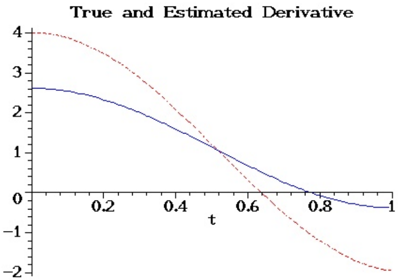
Example

The result with $\alpha = 10^{-2}$



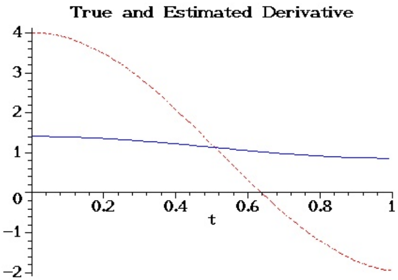
Example

The result with $\alpha = 0.1$



Example

The result with $\alpha = 1.0$



Choosing α

There are a variety of approaches to choosing α, all based on some analysis of the ill-posedness of the inverse problem and noise level in the data.

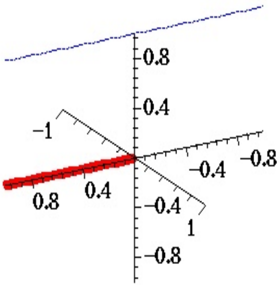
Choosing α

There are a variety of approaches to choosing α, all based on some analysis of the ill-posedness of the inverse problem and noise level in the data.

One practical approach to real problems is simulation to see what range for α works best.

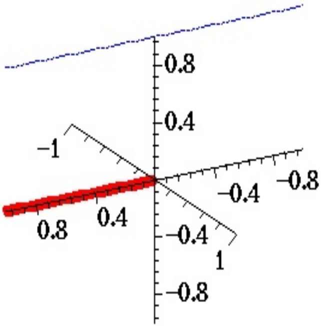
A 1D Version: The Forward Problem

A "1D" bar, length one meter, stretching along the x-axis from $0 < x < 1$ in 3D space, density $\lambda^*(x)$ Kg per meter of length. It has some gravitational field $\mathbf{F}$.



The Inverse Problem

Suppose we measure only the z component of the gravitational field, and only at points of the form $x_0 = t, y_0 = 0, z_0 = 1$.



The Inverse Problem

Can we find $\lambda^*(x)$ from $\mathbf{F}_z(t,0,1)$ and

$$\mathbf{F}_z(t,0,1) = \int_0^1 \frac{\lambda^*(x)}{((x-t)^2+1)^{3/2}}\,dx$$

The Inverse Problem

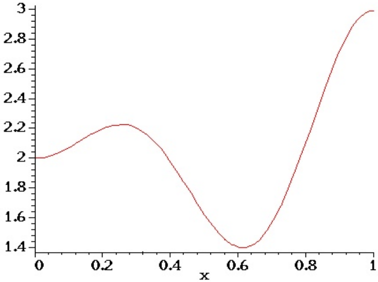
Can we find $\lambda^*(x)$ from $\mathbf{F}_z(t, 0, 1)$ and

$$\mathbf{F}_z(t, 0, 1) = \int_0^1 \frac{\lambda^*(x)}{((x-t)^2 + 1)^{3/2}}\, dx$$

One can prove this is possible, but the problem is unstable.

Example

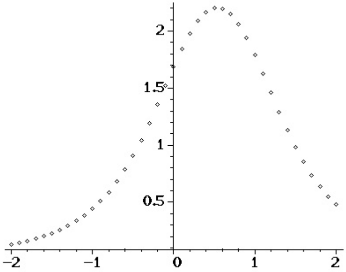
Suppose $\lambda^*(x) = 2 + x\sin(8x)$:



Let's try to fit $\lambda(x) = \sum_{k=0}^{10} a_k \cos(k\pi x)$ using least-squares.

Example

We take data at 40 points from $t = -2$ to $t = 2$:



Finding λ^*: Output Least Squares

If $\lambda^*(x)$ is the real density, let

$$h_k^* = \int_0^1 \frac{\lambda^*(x)}{((x - t_k)^2 + 1)^{3/2}}\, dx.$$

(So h_k^* is the real gravitational field at the point $(t_k, 0, 1)$).

Finding λ^*: Output Least Squares

If $\lambda^*(x)$ is the real density, let

$$h_k^* = \int_0^1 \frac{\lambda^*(x)}{((x-t_k)^2+1)^{3/2}}\,dx.$$

(So h_k^* is the real gravitational field at the point $(t_k, 0, 1)$).
Let

$$h_k(a_0, a_1, \ldots, a_{10}) = \int_0^1 \frac{\lambda(x)}{((x-t_k)^2+1)^{3/2}}\,dx.$$

We want to adjust $a_0, \ldots, a_{10}$ to make $h_k(a_0, \ldots, a_{10}) \approx h_k^*$, for all k.

Finding λ: Output Least Squares

Define the objective "fit-to-data" function

$$Q(a_0, \ldots, a_{10}) = \sum_{k=1}^{M} (h_k(a_0, \ldots, a_{10}) - h_k^*)^2.$$

Finding λ: Output Least Squares

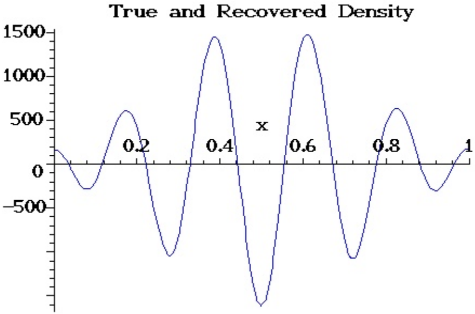
Define the objective "fit-to-data" function

$$Q(a_0, \ldots, a_{10}) = \sum_{k=1}^{M} (h_k(a_0, \ldots, a_{10}) - h_k^*)^2.$$

We settle for that $a_0, \ldots, a_{10}$ that minimizes Q, then take $\lambda(x) = \sum_{k=0}^{10} a_k \cos(k\pi x)$ as our estimate of the density.

Gravitational Prospecting Again

The result (no noise in the data) is



This inverse problem is severely ill-posed!

Regularizing The Problem

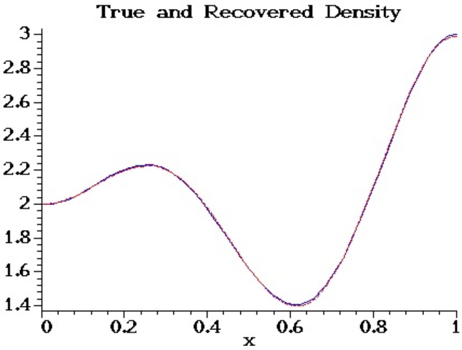
Add to Q a term that penalizes oscillatory behavior:

$$Q(a_0,\ldots,a_{10}) = \underbrace{\sum_{k=1}^{M}(h_k(a_0,\ldots,a_{10}) - h_k^*)^2}_{\text{fit to data}} + \underbrace{\alpha\int_0^1 (\lambda'(x))^2\,dx}_{\text{penalty term}}$$

$$= \underbrace{\sum_{k=1}^{M}(h_k(a_0,\ldots,a_{10}) - h_k^*)^2}_{\text{fit to data}} + \underbrace{\alpha\left(a_0^2 + \frac{\pi^2}{2}\sum_{k=1}^{10} k^2 a_k^2\right)}_{\text{penalty term}}.$$

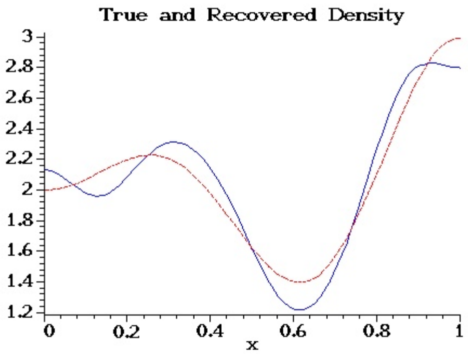
Regularization Result

The result (no noise in the data) is



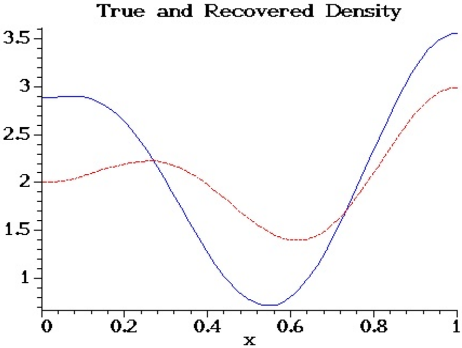
Regularization Result

Noise of magnitude 0.001 and $\alpha = 10^{-7}$:



Regularization Result

Noise of magnitude 0.1 and $\alpha = 10^{-4}$:



The General Framework for Tikhonov Regularization

We have a forward problem governed by an "operator" K:

$$K : \lambda \longrightarrow \text{observed data.}$$

The real parameter is λ^*, with real data d^*. We estimate λ^* by minimizing

$$Q(\lambda) = \|K(\lambda) - d^*\|^2 + \alpha P(\lambda)$$

where $P(\lambda)$ is a term that penalizes the behavior we don't want to see in λ (e.g., oscillation).

Notes URL

`www.rose-hulman.edu/~bryan/invprobs.html`