Estimation of a Non-Negative Function (Larkin, 1969)

Chris. J. Oates

July 25, 2017

10. Concluding Remarks

"... We have been concerned with the problem of estimating a function $f(r)$, about which we are given a limited amount of experimental information.

Our information about $f(r)$ is of two kinds:

1. General properties, such as non-negativity, continuity and integrability of $f(r)$ and/or its derivatives.

   This information, which may be genuinely given or arbitrarily imposed, limits *a priori* the class of allowable solutions. ("prior")

2. Specific properties, in the form of a finite number of observations on $f(r)$.

   These observations may be thought of as linear or non-linear functionals of f, which limit *a posteriori* the sub-class $\mathcal{C}$ to which f can belong. ("likelihood")

10. Concluding Remarks

"... We have been concerned with the problem of estimating a function $f(r)$, about which we are given a limited amount of experimental information.

Our information about $f(r)$ is of two kinds:

1. General properties, such as non-negativity, continuity and integrability of $f(r)$ and/or its derivatives.

   This information, which may be genuinely given or arbitrarily imposed, limits *a priori* the class of allowable solutions. ("prior")

2. Specific properties, in the form of a finite number of observations on $f(r)$.

   These observations may be thought of as linear or non-linear functionals of f, which limit *a posteriori* the sub-class $\mathcal{C}$ to which f can belong. ("likelihood")

10. Concluding Remarks

"... We have been concerned with the problem of estimating a function $f(r)$, about which we are given a limited amount of experimental information.

Our information about $f(r)$ is of two kinds:

1. General properties, such as non-negativity, continuity and integrability of $f(r)$ and/or its derivatives.

   This information, which may be genuinely given or arbitrarily imposed, limits *a priori* the class of allowable solutions. ("prior")

2. Specific properties, in the form of a finite number of observations on $f(r)$.

   These observations may be thought of as linear or non-linear functionals of f, which limit *a posteriori* the sub-class $\mathcal{C}$ to which f can belong. ("likelihood")

10. Concluding Remarks

"... We have been concerned with the problem of estimating a function $f(r)$, about which we are given a limited amount of experimental information.

Our information about $f(r)$ is of two kinds:

1. General properties, such as non-negativity, continuity and integrability of $f(r)$ and/or its derivatives.

   This information, which may be genuinely given or arbitrarily imposed, limits *a priori* the class of allowable solutions. ("prior")

2. Specific properties, in the form of a finite number of observations on $f(r)$.

   These observations may be thought of as linear or non-linear functionals of f, which limit *a posteriori* the sub-class $\mathcal{C}$ to which f can belong. ("likelihood")

10. Concluding Remarks

In general, an infinite number of exact observations will be required in order to characterise a particular $f \in \mathcal{C}$, an impossible requirement in a real, physical situation.

Thus, unless more assumptions are made, we can say no more about $f(r)$ than that it is a member of that sub-class of $\mathcal{C}$ whose elements could have resulted in the given observations. ("ill-posed")

A way out of this unpalatable but incontravertible dilemma is to express precisely any intuitive feeling we may have that certain members of $\mathcal{C}$ are *a priori* more likely than others. ("regularisation via a prior")

This enables us to formulate a variational problem whose solution, roughly speaking, represents the *a posteriori* most likely member of $\mathcal{C}$. ..." ("MAP estimation")

10. Concluding Remarks

In general, an infinite number of exact observations will be required in order to characterise a particular $f \in \mathcal{C}$, an impossible requirement in a real, physical situation.

Thus, unless more assumptions are made, we can say no more about $f(r)$ than that it is a member of that sub-class of $\mathcal{C}$ whose elements could have resulted in the given observations. ("ill-posed")

A way out of this unpalatable but incontravertible dilemma is to express precisely any intuitive feeling we may have that certain members of $\mathcal{C}$ are *a priori* more likely than others. ("regularisation via a prior")

This enables us to formulate a variational problem whose solution, roughly speaking, represents the *a posteriori* most likely member of $\mathcal{C}$. ..." ("MAP estimation")

10. Concluding Remarks

In general, an infinite number of exact observations will be required in order to characterise a particular $f \in \mathcal{C}$, an impossible requirement in a real, physical situation.

Thus, unless more assumptions are made, we can say no more about $f(r)$ than that it is a member of that sub-class of $\mathcal{C}$ whose elements could have resulted in the given observations. ("ill-posed")

A way out of this unpalatable but incontravertible dilemma is to express precisely any intuitive feeling we may have that certain members of $\mathcal{C}$ are *a priori* more likely than others. ("regularisation via a prior")

This enables us to formulate a variational problem whose solution, roughly speaking, represents the *a posteriori* most likely member of $\mathcal{C}$. ..." ("MAP estimation")

10. Concluding Remarks

In general, an infinite number of exact observations will be required in order to characterise a particular $f \in \mathcal{C}$, an impossible requirement in a real, physical situation.

Thus, unless more assumptions are made, we can say no more about $f(r)$ than that it is a member of that sub-class of $\mathcal{C}$ whose elements could have resulted in the given observations. ("ill-posed")

A way out of this unpalatable but incontravertible dilemma is to express precisely any intuitive feeling we may have that certain members of $\mathcal{C}$ are *a priori* more likely than others. ("regularisation via a prior")

This enables us to formulate a variational problem whose solution, roughly speaking, represents the *a posteriori* most likely member of $\mathcal{C}$. ..." ("MAP estimation")

Context

- Tikhonov, Andrey Nikolayevich (1943). On the stability of inverse problems (in Russian). Doklady Akademii Nauk SSSR. 39 (5): 195-198.
- Hoerl AE (1962), Application of ridge analysis to regression problems, Chemical Engineering Progress, 1958, 54-59
- Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method (translated in English). Soviet Mathematics. 4: 1035-1038.
- Larkin - paper submitted in 1968.

NB: Not ProbNum, just Inv Prob.

Context

- Tikhonov, Andrey Nikolayevich (1943). On the stability of inverse problems (in Russian). Doklady Akademii Nauk SSSR. 39 (5): 195-198.
- Hoerl AE (1962), Application of ridge analysis to regression problems, Chemical Engineering Progress, 1958, 54-59
- Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method (translated in English). Soviet Mathematics. 4: 1035-1038.
- Larkin - paper submitted in 1968.

NB: Not ProbNum, just Inv Prob.

2. The Particular Example

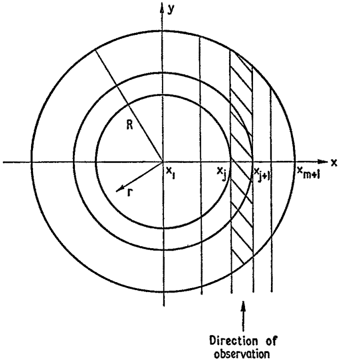

Fig. 1a. Diagram of the experimental situation.

Radiating cylinder of plasma

Observe the quantities

$$\varphi_j = 2 \int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r \mathrm{d}r$$

for $j = 1, \ldots, m$.

Task: Recover f. ("ill-posed")

2. The Particular Example

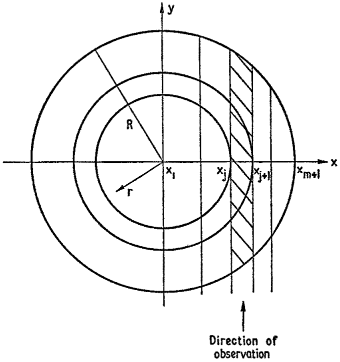

Fig. 1a. Diagram of the experimental situation.

Radiating cylinder of plasma

Observe the quantities

$$\varphi_j = 2\int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r\,\mathrm{d}r$$

for $j = 1, \ldots, m$.

Task: Recover f. ("ill-posed")

2. The Particular Example

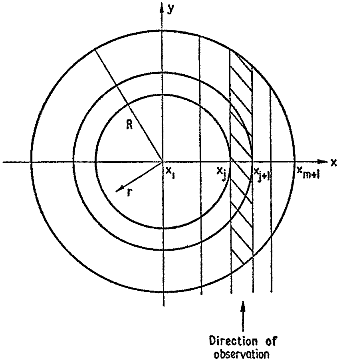

Fig. 1a. Diagram of the experimental situation.

Radiating cylinder of plasma

Observe the quantities

$$\varphi_j = 2 \int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r \mathrm{d}r$$

for $j = 1, \ldots, m$.

Task: Recover f. ("ill-posed")

2. The Particular Example

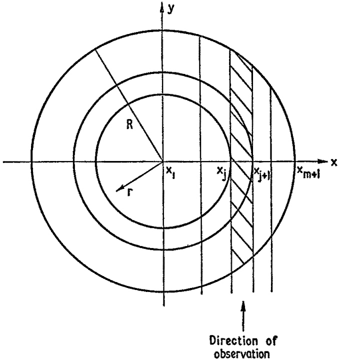

Fig. 1a. Diagram of the experimental situation.

Radiating cylinder of plasma

Observe the quantities

$$\varphi_j = 2 \int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r \mathrm{d}r$$

for $j = 1, \ldots, m$.

Task: Recover f. ("ill-posed")

5. A Relative Likelihood for Functions

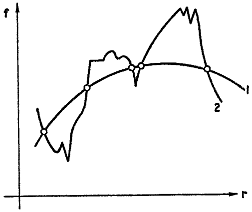

Fig. 2. Two functions which could have resulted in the same observed data points.

Entropy of radiation dist.

The relative likelihood

$$\mathcal{L}(f) := \int_0^R f(r) \log f(r) \, r \mathrm{d}r$$

("prior") leads to a variational problem

$$\arg\min \mathcal{L}(f) \quad \text{s.t.}$$
$$\varphi_j = 2 \int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r \mathrm{d}r$$

("MAP estimation")

Can prove that the argmin is non-negative (but fiddly).

5. A Relative Likelihood for Functions

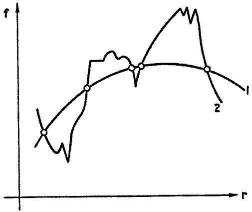

Fig. 2. Two functions which could have resulted in the same observed data points.

Entropy of radiation dist.

The relative likelihood

$$\mathcal{L}(f) := \int_0^R f(r) \log f(r) \; r\mathrm{d}r$$

("prior") leads to a variational problem

$$\arg\min \mathcal{L}(f) \quad \text{s.t.}$$

$$\varphi_j = 2 \int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r\mathrm{d}r$$

("MAP estimation")

Can prove that the argmin is non-negative (but fiddly).

5. A Relative Likelihood for Functions

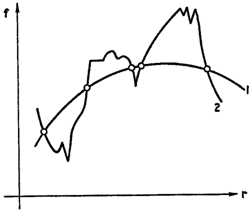

Fig. 2. Two functions which could have resulted in the same observed data points.

Entropy of radiation dist.

The relative likelihood

$$\mathcal{L}(f) := \int_0^R f(r) \log f(r)\ r\mathrm{d}r$$

("prior") leads to a variational problem

$$\arg\min \mathcal{L}(f) \quad \text{s.t.}$$

$$\varphi_j = 2 \int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r\mathrm{d}r$$

("MAP estimation")

Can prove that the argmin is non-negative (but fiddly).

5. A Relative Likelihood for Functions

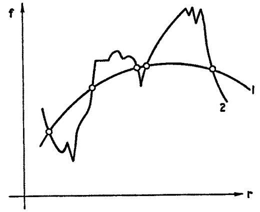

Fig. 2. Two functions which could have resulted in the same observed data points.

Entropy of radiation dist.

The relative likelihood

$$\mathcal{L}(f) := \int_0^R f(r) \log f(r) \; r\mathrm{d}r$$

("prior") leads to a variational problem

$$\arg\min \mathcal{L}(f) \quad \text{s.t.}$$

$$\varphi_j = 2 \int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r\mathrm{d}r$$

("MAP estimation")

Can prove that the argmin is non-negative (but fiddly).

5. A Relative Likelihood for Functions

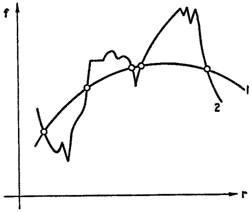

Fig. 2. Two functions which could have resulted in the same observed data points.

Entropy of radiation dist.

The relative likelihood

$$\mathcal{L}(f) := \int_0^R f(r) \log f(r)\; r\mathrm{d}r$$

("prior") leads to a variational problem

$$\arg\min \mathcal{L}(f) \quad \text{s.t.}$$

$$\varphi_j = 2 \int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r\mathrm{d}r$$

("MAP estimation")

Can prove that the argmin is non-negative (but fiddly).

5. A Relative Likelihood for Functions

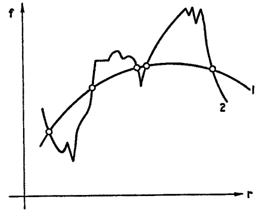

Fig. 2. Two functions which could have resulted in the same observed data points.

Entropy of radiation dist.

The relative likelihood

$$\mathcal{L}(f) := \int_0^R f(r) \log f(r)\; r\mathrm{d}r$$

("prior") leads to a variational problem

$$\arg\min \mathcal{L}(f) \quad \text{s.t.}$$

$$\varphi_j = 2 \int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r\mathrm{d}r$$

("MAP estimation")

Can prove that the argmin is non-negative (but fiddly).

5. A Relative Likelihood for Functions

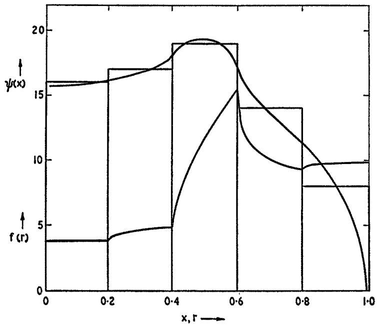

Fig. 3. Results from the 1st formulation.

But, whilst the Abel transform

$$\psi(x) = 2\int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r\,\mathrm{d}r$$

recovered in this way is smooth, the function $f(r)$ is *not* - against physical considerations.

5. A Relative Likelihood for Functions

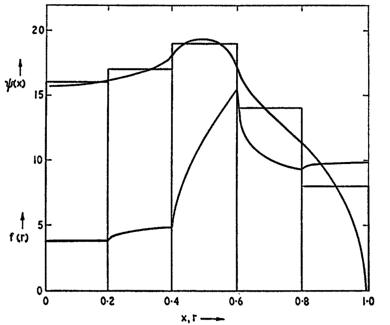

Fig. 3. Results from the 1st formulation.

But, whilst the Abel transform

$$\psi(x) = 2\int_x^R \frac{f(r)}{\sqrt{r^2 - x^2}} r \mathrm{d}r$$

recovered in this way is smooth, the function $f(r)$ is *not* - against physical considerations.

7. The Second Formulation

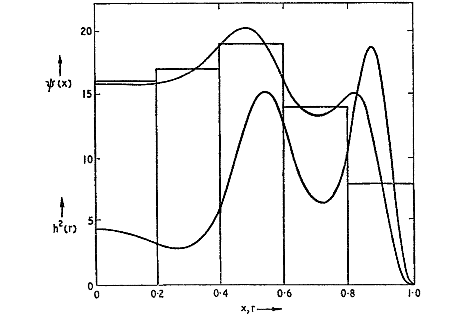

Fig. 4. Results from 2nd formulation, after 5 iterations; observations not smoothed.

Square root transform:

$$f(r) = h(r)^2$$

The relative likelihood

$$\mathcal{L}(h) := \int_0^R \left(\frac{\mathrm{d}h}{\mathrm{d}r}\right)^2 \, r\mathrm{d}r$$

("prior") leads to a variational problem

$$\arg\min \mathcal{L}(h) \quad \text{s.t.}$$

$$\varphi_j = 2\int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{h(r)^2}{\sqrt{r^2 - x^2}} r\mathrm{d}r$$

("MAP estimation")

Whilst the main contribution was MAP estimation, there is a hint toward Larkin 1972.

Namely, if h is considered random then the integrals

$$2\int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{h(r)^2}{\sqrt{r^2 - x^2}} r \mathrm{d}r$$

are also random - their distribution is derived in Section 8, but relies on the mathematics (Gaussian measures on Hilbert spaces) that was in preparation in Larkin 1972...

8. Functional Integration

Whilst the main contribution was MAP estimation, there is a hint toward Larkin 1972.

Namely, if h is considered random then the integrals

$$2\int_{x_j}^{x_{j+1}} \mathrm{d}x \int_x^R \frac{h(r)^2}{\sqrt{r^2 - x^2}} r \mathrm{d}r$$

are also random - their distribution is derived in Section 8, but relies on the mathematics (Gaussian measures on Hilbert spaces) that was in preparation in Larkin 1972...