Robust Inversion, Dimensionality Reduction, and Randomized Sampling

By Aleksandr Aravkin, Michael P. Friedlander, etc.

Shan You, Kai Zheng, Cong Fang

School of Information Science and Technology

December 15, 2014

Contents

Introduction

Consider the generic parameter-estimation scheme :

$$d_i = F_i(x)q_i + \epsilon_i \quad for \quad i = 1, .., m,$$

- observation d_i is obtained by the linear action of the forward model $F_i(x)$ on known source parameters q_i.
- ϵ_i captures the discrepancy between d_i and prediction $F_i(x)q_i$.

FWI Application in Seismology

This paper focuses the full-waveform inversion(FWI) application in seismology, which is used to image the earth's subsurface.

- forward model F : the solution operator of the wave equ.
- x : sound-velocity parameters for a 2- or 3-dimensional mesh.
- q_i encode the location and signature; d_i corresponding measurements.

Mathematical Methods

Minimizing some measure of misfit:

$$\min_{x} \quad \phi(x) := \frac{1}{m}\sum_{i=1}^{m}\phi_i(x)$$

where each $\phi_i(x)$ is some measure of the residual

$$r_i(x) := d_i - F_i(x)q_i$$

Typical Penalties

- Least-squares penalty : $\phi_i(x) = \|r_i(x)\|^2$
  equivalent to MAP estimate of x;
  ϵ_i independent and Gaussian.

- general ML or MAP estimation : $\phi_i(x) = -\log p_i(r_i(x))$
  where p_i is a particular probability density function of ϵ_i.

Goal & Main Work

Main Work:

Goal & Main Work

Main Work:

1. **Penalty Construction**
   Overcome the data contamination;
   From a statistical perspective: Student's t-distribution

Goal & Main Work

Main Work:

1. **Penalty Construction**
   Overcome the data contamination;
   From a statistical perspective: Student's t-distribution

2. **Dimensionality Reduction Technique**
   Costly computation of seismic data;
   Sample average approximations;
   Stochastic optimization

Basic Penalty Form

Natural Option for ML or MAP selection:

- using a log-concave density, $p(r) \propto \exp(-\rho(r))$, ρ convex penalty;
- $\phi_i(x) = \rho(r_i(x))$, for $i = 1, ..., m$.

Basic Penalty Form

Natural Option for ML or MAP selection:

- using a log-concave density, $p(r) \propto \exp(-\rho(r))$, ρ convex penalty;
- $\phi_i(x) = \rho(r_i(x))$, for $i = 1, ..., m$.

Two NOTES:

- for nonlinear F_i, typically nonconvex even for convex ρ;
- even for linear F_i, beneficial to choose a nonconvex ρ for outliers in the data.

Student's t-distribution

Student's t-density function:

$$p(r|\mu, \nu) \propto (1 + (r - \mu)^2/\nu)^{-(1+\nu)/2}$$

- heavy tail;
- the corresponding penalty function ($\mu = 0$): nonconvex

$$\rho(r) = \log(1 + r^2/\nu)$$

Outlier Removal

Outlier Removal

Typical question:

Given that a scalar residual deviates from the mean by more than t, what is the probability that it actually deviates by more than $2t$?

Outlier Removal

Typical question:

Given that a scalar residual deviates from the mean by more than t, what is the probability that it actually deviates by more than $2t$?

- 1-norm (the slowest-growing convex penalty; Laplace distribution with mean $1/\alpha$)

$$Pr(|r| > t_2 || r| > t_1) = Pr(|r| > t_2 - t_1) = \exp(-\alpha[t_2 - t_1])$$

Outlier Removal

- Student's t-distribution (Cauchy distribution with $\nu = 1$)

$$\lim_{t \to \infty} Pr(|r| > 2t \,|\, |r| > t) = \lim_{t \to \infty} \frac{\frac{\pi}{2} - \arctan(2t)}{\frac{\pi}{2} - \arctan(t)} = \frac{1}{2}$$

- general convex penalty (differentiable; proved)

$$Pr(|r| > t_2 \,|\, |r| > t_1) = Pr(|r| > t_2 - t_1) \leqslant \exp(-\alpha_0[t_2 - t_1])$$

Outlier Removal

- Student's t-distribution (Cauchy distribution with $\nu = 1$)

$$\lim_{t\to\infty} Pr(|r| > 2t||r| > t) = \lim_{t\to\infty} \frac{\frac{\pi}{2} - \arctan(2t)}{\frac{\pi}{2} - \arctan(t)} = \frac{1}{2}$$

- general convex penalty (differentiable; proved)

$$Pr(|r| > t_2||r| > t_1) = Pr(|r| > t_2 - t_1) \leqslant \exp(-\alpha_0[t_2 - t_1])$$

One critical conclusion:

Log-concave density family 'ignores' the existence of outliers to some extent while the Student's t-distribution doesn't.

Outlier Removal

Another perspective: **Influence Function** $\rho'(t)$

- Laplace: sign function; Gaussian: linear function
- Student's t-density:

$$\rho'(r) = \frac{2r}{\nu + r^2}$$

Outlier Removal

Another perspective: **Influence Function** $\rho'(t)$

- Laplace: sign function; Gaussian: linear function
- Student's t-density:

$$\rho'(r) = \frac{2r}{\nu + r^2}$$

Tradeoff:

- Convex models are easier to characterize and solve, but may be wrong in a situation in which large outliers are expected.
- Nonconvex penalties are particularly useful with large outliers.

Explicit Diagram

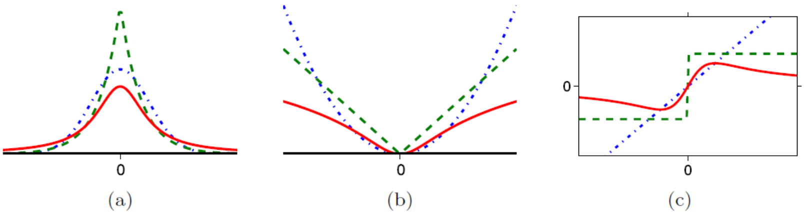

Fig. 1: The Gaussian (·–), Laplace (– –), and Student's t- (—) distributions: (a) densities, (b) penalties, and (c) influence functions.

Thank you!