

Numerical
Differentiation

Anders
Munk-Nielsen

Introduction

Estimating Noise
Level

Black Box
Differentiation

Numerical Differentiation

NUMEDIG Seminar

Anders Munk-Nielsen
Department of Economics



Numerical
Differentiation

Anders
Munk-Nielsen

Agenda

Disclaimer: These slides borrow heavily from slides and presentation by Stefan Wild and ZICE 2014.

Introduction

Estimating Noise
Level

Black Box
Differentiation

① Introduction

② Estimating Noise Level

③ Black Box Differentiation



Numerical
Differentiation

Anders
Munk-Nielsen

Outline

Introduction

Estimating Noise
Level

Black Box
Differentiation

① Introduction

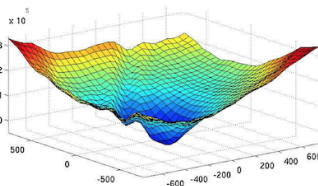
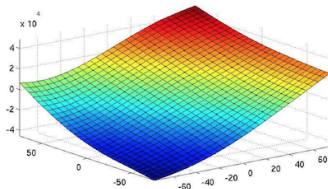
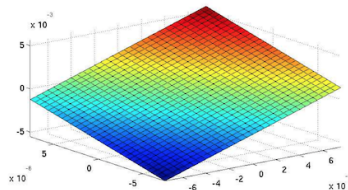
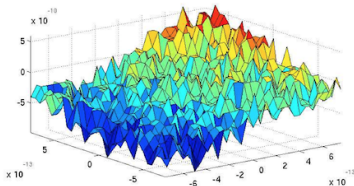
② Estimating Noise Level

③ Black Box Differentiation



The Problem

Here are 4 pictures of the same function (varying the zoom)



[picture shamelessly borrowed from Stefan Wild's slides]



The Idea — Finite Differences (FD)

- **Problem:** For $f : \mathbb{R}^K \rightarrow \mathbb{R}$, approximate the entries of $\nabla f(x)$, namely

$$\frac{df(x)}{dx_k} = \lim_{h \searrow 0} \frac{f(x_1, \dots, x_k + h, \dots, x_K) - f(x)}{h}.$$

- **Basic method:** Finite difference (FD) takes

$$\frac{df(x)}{dx_k} \cong \frac{\Delta f(x)}{\Delta x_k},$$

where $\Delta x_k = h > 0$ is some small number.

- **Problem:** How to choose h .
- **Trade-off:**
 - Truncation error
(because $h \neq 0$)
 - Roundoff error
(because of finite precision)
- [show in Matlab for known, analytic functions]



Numerical
DifferentiationAnders
Munk-Nielsen

Introduction

Estimating Noise
LevelBlack Box
Differentiation

Why care?

- **Expensive objective:** Then you want to economize on the number of $f(\theta)$ evaluations.
 - Analytical derivatives
 - Black box differentiation
 - Control the Hessian approximation (e.g. BHHH)
- **Inexpensive:** Then you want to be as precise as possible.
 - Use central differences — $O(h^2)$ instead of $O(h)$,

$$\frac{df(x)}{dx_k} \cong \frac{f(x+h) - f(x-h)}{2h} = O(h^2).$$

- Consider computing analytic or full FD Hessian.



Numerical
Differentiation

Anders
Munk-Nielsen

Outline

Introduction

Estimating Noise
Level

Black Box
Differentiation

① Introduction

② Estimating Noise Level

③ Black Box Differentiation



Noise Level

- **Simple model:**

$$f(x) = f^*(x) + \varepsilon(x), \quad x \in \mathcal{I}, \varepsilon(x) \sim \text{IID}.$$

- **Pedantic:** In a *really* small neighbourhood, \mathcal{I} , x are definitely systematically related.

- **Definition:** The noise level of f is

$$\varepsilon_f^2 \equiv \sqrt{\text{Var} [\varepsilon(x)]}.$$

- **Step:** The optimal step is $\cong \sqrt{\varepsilon_f^2}$.

- Because

$$\begin{aligned} \text{error} &= \text{truncation} + \text{roundoff} = \frac{h}{2}M + \frac{\varepsilon_f^2 B}{h} \\ \Rightarrow \text{FOC} : 0 &= \frac{M}{2} - \varepsilon_f^2 B h^{-2} \\ \Leftrightarrow h^* &= \sqrt{\varepsilon_f^2 \frac{2B}{M}}. \end{aligned}$$



Algorithm: ECNoise

- **Idea:** For smooth f , $f^{(k)}(x) \rightarrow 0$ “quickly”.
- **Define:** Finite difference,

$$\Delta^{k+1}f(x) = \Delta^k f(x+h) - \Delta^k f(x), \quad \Delta^0 f(x) = f(x).$$

- **Idea:** If h is small enough, $\Delta^k f(x) \cong \Delta^k \varepsilon(x)$.
- **Why?** If f^* is k times differentiable,

$$\Delta^k f(x) = (f^*)^{(k)}(\xi_k)h^k + \Delta^k \varepsilon(x), \quad \xi \in (x; x+kh).$$

- \Rightarrow **Goal:** Make h small enough to remove the smooth part.



Theory

- **Let:** $\varepsilon(x) \sim \text{IID}(0, \varepsilon_f^2)$

① $\mathbb{E} \left[\Delta^k \varepsilon(x) \right] = 0.$

② $\gamma_k \mathbb{E} \left\{ \left[\Delta^k \varepsilon(x) \right]^2 \right\} = \text{Var}(\varepsilon) \equiv \varepsilon_f^2, \text{ for } \gamma_k \equiv \frac{(k!)^2}{(2k)!}.$

- ③ If f^* is continuous at x , then

$$\lim_{h \searrow 0} \gamma_k \mathbb{E} \left\{ \left[\Delta^k f(x) \right]^2 \right\} = \varepsilon_f^2.$$

- ④ If f^* is k -times continuously differentiable at x , then

$$\lim_{h \searrow 0} \frac{\gamma_k \mathbb{E} \left\{ \left[\Delta^k f(x) \right]^2 \right\} - \varepsilon_f^2}{h^{2k}} = \gamma_k \left[(f^*)^{(k)}(x) \right]^2.$$

- **Then:** For h sufficiently small,

$$\Rightarrow \varepsilon_f^2 \cong \gamma_k \mathbb{E} \left\{ \left[\Delta^k f(x) \right]^2 \right\}.$$



Numerical
DifferentiationAnders
Munk-Nielsen

Introduction

Estimating Noise
LevelBlack Box
Differentiation

Multivariate?

- **Idea:** For $f : \mathbb{R}^K \rightarrow \mathbb{R}$, simply choose a basis point and a direction, $x_0, p \in \mathbb{R}^K$, and apply the theory from above on

$$\tilde{f}(x) = f(x_0 + xp), \quad x \geq 0.$$

- **Directional:** Use directional derivatives and differences.
- **Warning:** The estimated noise level, σ_ε , may now depend on the chosen p .
- **Solution:** Use a random direction, p .



Numerical
Differentiation

Anders
Munk-Nielsen

Outline

Introduction

Estimating Noise
Level

Black Box
Differentiation

① Introduction

② Estimating Noise Level

③ Black Box Differentiation



Motivation

- **Econometrics:** Often the criterion function has the form,

$$f(\theta) = \sum_{i=1}^N \log l(\theta, z_i),$$

$$f(\theta) = \sum_{i=1}^N \varphi \left(\left\| y_i^{\text{pred}}(\theta) - y_i^{\text{obs}} \right\| \right),$$

$$f(\theta) = \sum_{i=1}^N q(\theta, z_i),$$

where $l_i(\cdot, \cdot)$, $\varphi(\cdot)$ and $q(\cdot, \cdot)$ are well-known **black box functions**.

- **Goal:** Prefer not to differentiate l .
- **Idea:** Utilize the structure we *do* know:

$$\nabla f(\theta) = \sum_{i=1}^N \frac{1}{l(\theta, z_i)} \nabla l(\theta, z_i).$$



Numerical
Differentiation

Anders
Munk-Nielsen

Introduction

Estimating Noise
Level

Black Box
Differentiation

Why?

- Why does it help?
 - ① We use more structure,
(so we can't be worse off in terms of roundoff)
 - ② Better Hessian approximation (BHHH or Taylor approximation),

$$H \cong \nabla f(\theta)' \nabla f(\theta)$$

- **Note:** It doesn't help (much) for the gradient.



Other comments

- **Ingredients:** An optimization algorithm must have two ingredients
 - 1 Globalization strategy (always converge to *something*),
 - 1 Line-search
 - 2 Trust region
(only available in `fminunc` when using analytic gradients)
 - 2 Local method (how to find the step size),
 - 1 Sequential quadratic and/or linear programming
 - 2 Interior-point
- **Ad 1)** Avoid cycles, avoid overstepping.
- **Ad 2)** Choosing the best step given the problem at hand, gradient and hessian.



Logit Example

- **Model:** N consumers are faced with J different cars, having K characteristics.
- **Random utility:** Car j yields utility

$$u_{ij} = \beta_i x_j, \quad \beta_i = z_i \beta.$$

- **Likelihood:**

$$\Pr(d|z_i; \theta) = \frac{\exp(z_i \beta x_j)}{\sum_{j'=1}^J \exp(z_i \beta x_{j'})}.$$

- **Derivative:**

$$\nabla_{\theta} \log \Pr(d|z_i; \theta) = \nabla_{\theta} u_{id} - \sum_{j'=1}^J \Pr(j'|z_i; \theta) \nabla_{\theta} u_{ij'}.$$

- **BHHH:** For the Hessian, use the approximation

$$H \cong -N^{-1} g' g, \quad g = \begin{pmatrix} \nabla_{\theta} \log \Pr(d_1|z_1; \theta) \\ \vdots \\ \nabla_{\theta} \log \Pr(d_N|z_N; \theta) \end{pmatrix}.$$

- **Finding:** Trust-region helped *a lot* for the logit example.



Numerical
Differentiation

Anders
Munk-Nielsen

Introduction

Estimating Noise
Level

Black Box
Differentiation

Setting

- **Gradient:** Numeric, analytic or black box.
- **Hessian:** BFGS-updating or BHHH.
- **Local method:**
 - Quasi-Newton/line search — only available method with numeric gradients.
 - Trust-region — Much more efficient.



Numerical
Differentiation

Anders
Munk-Nielsen

The End

The end

Introduction

Estimating Noise
Level

Black Box
Differentiation

