Numerical Computation of Derivatives

John Bagterp Jørgensen

Department of Applied Mathematics and Computer Science
Technical University of Denmark

02612 Constrained Optimization

Outline

Topic: How to compute derivatives numerically.

Read Chap. 8 in Nocedal & Wright and Appendix A of the Lecture Notes

Univariate Function, $x \in \mathbb{R}$ (scalar function)

- Argument: $x \in \mathbb{R}$
- Function: $f = f(x)$, $f : \mathbb{R} \mapsto \mathbb{R}$
- Gradient: $g = g(x) = \nabla f(x) = \frac{df}{dx}(x)$, $g : \mathbb{R} \mapsto \mathbb{R}$
- Hessian: $H = H(x) = \nabla^2 f(x) = \frac{d^2 f}{dx^2}(x)$, $H : \mathbb{R} \mapsto \mathbb{R}$
  Note that: $H = \nabla(\nabla f(x)) = \nabla g(x) = \frac{d}{dx}\left(\frac{df}{dx}(x)\right) = \frac{dg}{dx}(x)$

Finite difference approximations of the gradient

- Forward Difference (FD): $g(x) = \frac{df}{dx}(x) \approx \frac{f(x+h)-f(x)}{h}$
- Backward Difference (BD): $g(x) = \frac{df}{dx}(x) \approx \frac{f(x)-f(x-h)}{h}$
- Central Difference (CD): $g(x) = \frac{df}{dx}(x) \approx \frac{f(x+h)-f(x-h)}{2h}$

Finite difference approximations of the Hessian

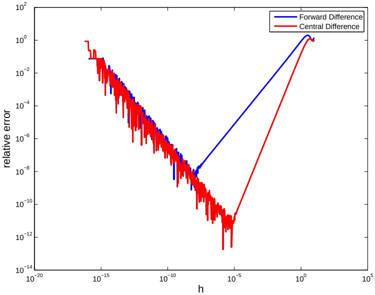
- FD from gradient: $H(x) = \frac{dg}{dx}(x) \approx \frac{g(x+h)-g(x)}{h}$
- BD from gradient: $H(x) = \frac{dg}{dx}(x) \approx \frac{g(x)-g(x-h)}{h}$
- CD from gradient: $H(x) = \frac{dg}{dx}(x) \approx \frac{g(x+h)-g(x-h)}{2h}$
- Hessian from function: $H(x) = \frac{d^2 f}{dx^2}(x) \approx \frac{f(x+h)-2f(x)+f(x-h)}{h^2}$

Evaluation of f and g from the function $f(x)$
Effect of step size, h

Numerical example function:

- Point: $x = 1.0$
- Function: $f(x) = \sin(x)$
- Gradient: $g(x) = \cos(x)$
- Hessian: $H(x) = -\sin(x)$

Machine precision:
$u = 2.2 \cdot 10^{-16}$, $u^{1/2} = 1.5 \cdot 10^{-8}$, $u^{1/3} = 6.1 \cdot 10^{-6}$



Finite difference gradient approximation

- Forward Difference (FD):

$$g_{FD} = g_{FD}(x; h) = \frac{f(x+h) - f(x)}{h}$$

- Central Difference (CD):

$$g_{CD} = g_{CD}(x; h) = \frac{f(x+h) - f(x-h)}{h}$$

- Error

$$e_{FD} = \|g_{FD}(x; h) - g(x)\|$$
$$e_{CD} = \|g_{CD}(x; h) - g(x)\|$$

- Relative error

$$\varepsilon_{FD} = \frac{e_{FD}}{\max\{1.0, \|g(x)\|\}}$$
$$\varepsilon_{CD} = \frac{e_{CD}}{\max\{1.0, \|g(x)\|\}}$$

Round-off Error - Finite Precision Floating Point Arithmetics

- Floating-point function evaluation ($\tilde{f}$), exact function evaluation (f), and rounding error (e):

$$
\begin{aligned}
\tilde{f}(x) &= f(x) + e_k & \qquad \|e_k\| &\leq u \\
\tilde{f}(x+h) &= f(x+h) + e_{k+1} & \qquad \|e_{k+1}\| &\leq u \\
\tilde{f}(x-h) &= f(x-h) + e_{k-1} & \qquad \|e_{k-1}\| &\leq u
\end{aligned}
$$

  Machine precision: $u = 2.22 \cdot 10^{-16}$
- Finite-difference gradient approximation and round-off error

$$
\begin{aligned}
\tilde{g}_{FD} &= \frac{\tilde{f}(x+h) - \tilde{f}(x)}{h} = \frac{f(x+h) - f(x)}{h} + \frac{e_{k+1} - e_k}{h} = g_{FD} + r_{FD} \\
\tilde{g}_{BD} &= \frac{\tilde{f}(x) - \tilde{f}(x-h)}{h} = \frac{f(x) - f(x-h)}{h} + \frac{e_k - e_{k-1}}{h} = g_{BD} + r_{BD} \\
\tilde{g}_{CD} &= \frac{\tilde{f}(x+h) - \tilde{f}(x-h)}{2h} = \frac{f(x+h) - f(x-h)}{2h} + \frac{e_{k+1} - e_{k-1}}{2h} = g_{CD} + r_{CD}
\end{aligned}
$$

- Round-off error for finite-difference gradient approximation

$$
\begin{aligned}
\|r_{FD}\| &\leq \frac{2u}{h} & \qquad \log\|r_{FD}\| &\leq \log(2u) - \log(h) \\
\|r_{BD}\| &\leq \frac{2u}{h} & \qquad \log\|r_{BD}\| &\leq \log(2u) - \log(h) \\
\|r_{CD}\| &\leq \frac{u}{h} & \qquad \log\|r_{CD}\| &\leq \log(u) - \log(h)
\end{aligned}
$$

Asymptotic Analysis and the Taylor Approximation - FD

- Taylor approximation

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \frac{1}{6}f'''(x)h^3 + O(h^4)$$

- Forward difference gradient approximation error

$$\begin{aligned}
g_{FD} &= \frac{f(x+h) - f(x)}{h} = f'(x) + \frac{1}{2}f''(x)h + \frac{1}{6}h^2 + O(h^3) \\
e_{FD} &= \|g_{FD}(x;h) - g(x)\| \\
&\approx \left\| f'(x) + \frac{1}{2}f''(x)h + \frac{1}{6}f'''(x)h^2 - f'(x) \right\| \\
&= \left\| \frac{1}{2}f''(x)h + \frac{1}{6}f'''(x)h^2 \right\| \approx \left\| \frac{1}{2}f''(x) \right\| h = \alpha h
\end{aligned}$$

- Total error

$$\epsilon = e_{FD} + r_{FD} \approx \alpha h + \frac{2u}{h}$$

Optimal total error: $\frac{d\epsilon}{dh} = \alpha - (2u)h^{-2} = 0$ such that $h = \sqrt{\frac{2u}{\alpha}} = \sqrt{\frac{4u}{\|f''(x)\|}}$

Asymptotic Analysis and the Taylor Approximation - BD

- Taylor approximation

$$f(x-h) = f(x) - f'(x)h + \frac{1}{2}f''(x)h^2 - \frac{1}{6}f'''(x)h^3 + O(h^4)$$

- Backward difference gradient approximation error

$$\begin{aligned}
g_{BD} &= \frac{f(x) - f(x-h)}{h} = f'(x) - \frac{1}{2}f''(x)h + \frac{1}{6}f'''(x)h^2 + O(h^3) \\
e_{BD} &= \|g_{BD}(x;h) - g(x)\| \\
&\approx \left\| f'(x) - \frac{1}{2}f''(x)h + \frac{1}{6}f'''(x)h^2 - f'(x) \right\| \\
&= \left\| -\frac{1}{2}f''(x)h + \frac{1}{6}f'''(x)h^2 \right\| \approx \left\| \frac{1}{2}f''(x) \right\| h = \alpha h
\end{aligned}$$

- Total error

$$\epsilon = e_{BD} + r_{BD} \approx \alpha h + \frac{2u}{h}$$

Optimal total error: $\frac{d\epsilon}{dh} = \alpha - (2u)h^{-2} = 0$ such that $h = \sqrt{\frac{2u}{\alpha}} = \sqrt{\frac{4u}{\|f''(x)\|}}$

Asymptotic Analysis and the Taylor Approximation - CD

- Taylor approximation

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \frac{1}{6}f'''(x)h^3 + \frac{1}{24}f''''(x)h^4 + O(h^5)$$
$$f(x-h) = f(x) - f'(x)h + \frac{1}{2}f''(x)h^2 - \frac{1}{6}f'''(x)h^3 + \frac{1}{24}f''''(x)h^4 + O(h^5)$$

- Central difference gradient approximation error

$$\begin{aligned} g_{CD} &= \frac{f(x+h) - f(x-h)}{2h} \approx f'(x) + \frac{1}{6}f'''(x)h^2 + O(h^4) \\ e_{CD} &= \|g_{CD}(x;h) - g(x)\| \\ &\approx \left\| f'(x) + \frac{1}{6}f'''(x)h^2 - f'(x) \right\| \\ &= \left\| \frac{1}{6}f'''(x)h^2 \right\| = \left\| \frac{1}{6}f'''(x) \right\| h^2 = \beta h^2 \end{aligned}$$

- Total error

$$\epsilon = e_{CD} + r_{CD} \approx \beta h^2 + \frac{u}{h}$$

Optimal total error:

$\frac{d\epsilon}{dh} = 2\beta h - uh^{-2} = 0$ such that $h = \left(\frac{u}{2\beta}\right)^{\frac{1}{3}} = \left(\frac{3u}{\|f'''(x)\|}\right)^{\frac{1}{3}}$

Evaluation of f, g, and H from the function $f(x)$

- Evaluate the objective function:

$$f = f_k = f(x)$$

- Evaluate the objective function in neighboring points:

$$f_{k+1} = f(x+h)$$
$$f_{k-1} = f(x-h)$$

- Gradient approximation:

$$g \approx \frac{f(x+h) - f(x-h)}{2h} = \frac{f_{k+1} - f_{k-1}}{2h}$$

- Hessian approximation:

$$H \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} = \frac{f_{k+1} - 2f_k + f_{k-1}}{h^2}$$

Bivariate Function, $x \in \mathbb{R}^2$

- Function: $f = f(x, y)$
- Finite difference gradient approximations

$$g_{FD} = \frac{num}{den}$$

- Function: $f = f(x, y)$
- Gradient approximation (central difference)

$$g = g(x, y) = \nabla f(x, y) = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$$

$$f_x(x, y) \approx \frac{f(x + \Delta x, y) - f(x - \Delta x, y)}{\Delta x}$$

$$f_y(x, y) \approx \frac{f(x, y + \Delta y) - f(x, y - \Delta y)}{\Delta y}$$

- Hessian approximation

$$H = H(x, y) = \nabla^2 f(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix}$$

$$f_{xx}(x, y) \approx \frac{f(x + \Delta x, y) - 2f(x, y) + f(x - \Delta x, y)}{\Delta x^2}$$

$$f_{yy}(x, y) \approx \frac{f(x, y + \Delta y) - 2f(x, y) + f(x, y - \Delta y)}{\Delta y^2}$$

$$f_{xy}(x, y) \approx \frac{f(x + \Delta x, y + \Delta y) - f(x + \Delta x, y - \Delta y) - f(x - \Delta x, y + \Delta y) + f(x - \Delta x, y - \Delta y)}{4\Delta x \Delta y}$$

$$f_{yx}(x, y) = f_{xy}(x, y)$$

- Function evaluations: $1 + 4 + 4 = 9$
  Function: $f(x, y)$
  Gradient: $f(x + \Delta x, y)$, $f(x - \Delta x, y)$, $f(x, y + \Delta y)$, $f(x, y - \Delta y)$
  Hessian: $f(x + \Delta x, y + \Delta y)$, $f(x + \Delta x, y - \Delta y)$, $f(x - \Delta x, y + \Delta y)$, $f(x - \Delta x, y - \Delta y)$

- Function: $f = f(x)$, $x \in \mathbb{R}^2$
- Gradient approximation (central difference)

$$g = g(x) = \nabla f(x) = \begin{bmatrix} f_{x_1}(x) \\ f_{x_2}(x) \end{bmatrix}$$

$$f_{x_1}(x) \approx \frac{f(x + e_1 \Delta x_1) - f(x - e_1 \Delta x_1)}{\Delta x_1} = \frac{f_{1+} - f_{1-}}{\Delta x_1}$$

$$f_{x_2}(x) \approx \frac{f(x + e_2 \Delta x_2) - f(x - e_2 \Delta x_2)}{\Delta x_2} = \frac{f_{2+} - f_{2-}}{\Delta x_2}$$

- Hessian approximation

$$H = H(x) = \nabla^2 f(x) = \begin{bmatrix} f_{x_1 x_1}(x) & f_{x_1 x_2}(x) \\ f_{x_2 x_1}(x) & f_{x_2 x_2}(x) \end{bmatrix}$$

$$f_{x_1 x_1}(x) \approx \frac{f(x + e_1 \Delta x_1) - 2f(x) + f(x - e_1 \Delta x_1)}{\Delta x_1^2} = \frac{f_{1+} - 2f + f_{1-}}{\Delta x_1^2}$$

$$f_{x_2 x_2}(x) \approx \frac{f(x + e_2 \Delta x_2) - 2f(x) + f(x - e_2 \Delta x_2)}{\Delta x_2^2} = \frac{f_{2+} - 2f + f_{2-}}{\Delta x_2^2}$$

$$f_{x_1 x_2}(x) \approx \frac{f_{++} - f_{+-} - f_{-+} + f_{--}}{4 \Delta x_1 \Delta x_2}$$

$$f_{x_2 x_1}(x) = f_{x_1 x_2}(x)$$

- Function evaluations: $1 + 4 + 4 = 9$
  Function: $f = f(x)$
  Gradient: $f_{1+} = f(x + e_1 \Delta x_1)$, $f_{1-} = f(x - e_1 \Delta x_1)$,
  $f_{2+} = f(x + e_2 \Delta x_2)$, $f_{2-} = f(x - e_2 \Delta x_2)$
  Hessian: $f_{++} = f(x + e_1 \Delta x_1 + e_2 \Delta x_2)$, $f_{+-} = f(x + e_1 \Delta x_1 - e_2 \Delta x_2)$,
  $f_{-+} = f(x - e_1 \Delta x_1 + e_2 \Delta x_2)$, $f_{--} = f(x - e_1 \Delta x_1 - e_2 \Delta x_2)$

Multivariate Function, $x \in \mathbb{R}^n$

Objective function, gradient, and the Hessian matrix

- Vector: $x \in \mathbb{R}^n$
- Objective function: $f : \mathbb{R}^n \mapsto \mathbb{R}$

$$f = f(x)$$

- Gradient: $g : \mathbb{R}^n \mapsto \mathbb{R}^n$

$$g = g(x) = \nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$$

- Hessian matrix: $H : \mathbb{R}^n \mapsto \mathbb{R}^{n \times n}$

$$H = H(x) = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right)(x) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right)(x) & \dots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_1}\right)(x) \\ \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right)(x) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right)(x) & \dots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_2}\right)(x) \\ \\ \vdots \\ \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_n}\right)(x) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_n}\right)(x) & \dots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_n}\right)(x) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1}\nabla f(x) & \frac{\partial}{\partial x_2}\nabla f(x) & \dots & \frac{\partial}{\partial x_n}\nabla f(x) \end{bmatrix}$$

Nonlinear Least Squares Objective Function

- Residual function and Jacobian

$$r(x) = \begin{bmatrix} r_1(x) \\ r_2(x) \\ \vdots \\ r_m(x) \end{bmatrix} \qquad J(x) = \begin{bmatrix} \frac{\partial r_1}{\partial x_1}(x) & \frac{\partial r_1}{\partial x_2}(x) & \dots & \frac{\partial r_1}{\partial x_n}(x) \\ \frac{\partial r_2}{\partial x_1}(x) & \frac{\partial r_2}{\partial x_2}(x) & \dots & \frac{\partial r_2}{\partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial r_m}{\partial x_1}(x) & \frac{\partial r_m}{\partial x_2}(x) & \dots & \frac{\partial r_m}{\partial x_n}(x) \end{bmatrix}$$

- Nonlinear least squares objective function

$$f = f(x) = \frac{1}{2} \|r(x)\|_2^2 = \frac{1}{2} r(x)' r(x) = \frac{1}{2} \sum_{k=1}^{m} r_k(x)^2$$

- Gradient

$$g = g(x) = \nabla f(x) = J(x)' r(x)$$

- Hessian

$$H = H(x) = \nabla^2 f(x) = J(x)' J(x) + \sum_{k=1}^{m} r_k(x) \nabla^2 r_k(x)$$

Nocedal and Wright p 245-246

Himmelblau's Objective Function

- Objective function

$$f = f(x) = \left(x_1^2 + x_2 - 11\right)^2 + \left(x_1 + x_2^2 - 7\right)^2 = \frac{1}{2} \|r(x)\|_2^2$$

- Residual functions

$$r(x) = \begin{bmatrix} r_1(x) \\ r_2(x) \end{bmatrix} = \begin{bmatrix} \sqrt{2}\left(x_1^2 + x_2 - 11\right) \\ \sqrt{2}\left(x_1 + x_2^2 - 7\right) \end{bmatrix}$$

- Jacobian

$$J(x) = \begin{bmatrix} \nabla r_1(x)' \\ \nabla r_2(x)' \end{bmatrix} = \begin{bmatrix} \frac{\partial r_1}{\partial x_1}(x) & \frac{\partial r_1}{\partial x_2}(x) \\ \frac{\partial r_2}{\partial x_1}(x) & \frac{\partial r_2}{\partial x_2}(x) \end{bmatrix} = \begin{bmatrix} 2\sqrt{2}x_1 & \sqrt{2} \\ \sqrt{2} & 2\sqrt{2}x_2 \end{bmatrix}$$

- Hessians of the residual functions

$$\nabla^2 r_1(x) = \begin{bmatrix} \frac{\partial^2 r_1}{\partial x_1^2}(x) & \frac{\partial^2 r_1}{\partial x_1 \partial x_2}(x) \\ \frac{\partial^2 r_1}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 r_1}{\partial x_2^2}(x) \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix}$$

$$\nabla^2 r_2(x) = \begin{bmatrix} \frac{\partial^2 r_2}{\partial x_1^2}(x) & \frac{\partial^2 r_2}{\partial x_1 \partial x_2}(x) \\ \frac{\partial^2 r_2}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 r_2}{\partial x_2^2}(x) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 2\sqrt{2} \end{bmatrix}$$

- Gradient and Hessian

$$g = g(x) = \nabla f(x) = J(x)' r(x)$$

$$H = H(x) = \nabla^2 f(x) = J(x)' J(x) + \sum_{k=1}^{m=2} r_k(x) \nabla^2 r_k(x)$$

- Function: $f = f(x)$, $x \in \mathbb{R}^n$
- Gradient approximation (central difference)

$$g = g(x) = \nabla f(x) = \begin{bmatrix} f_{x_1}(x) \\ f_{x_2}(x) \\ \vdots \\ f_{x_n}(x) \end{bmatrix}$$

$$f_{x_i}(x) \approx \frac{f(x + e_i \Delta x_i) - f(x - e_i \Delta x_i)}{\Delta x_i} = \frac{f_{i+} - f_{i-}}{\Delta x_i} \qquad i = 1, 2, \ldots, n$$

- Hessian approximation

$$H = H(x) = \nabla^2 f(x) = \begin{bmatrix} f_{x_1 x_1}(x) & f_{x_1 x_2}(x) & \ldots & f_{x_1 x_n}(x) \\ f_{x_2 x_1}(x) & f_{x_2 x_2}(x) & \ldots & f_{x_2 x_n}(x) \\ \vdots & \vdots & & \vdots \\ f_{x_n x_1}(x) & f_{x_n x_2}(x) & \ldots & f_{x_n x_n}(x) \end{bmatrix}$$

$$f_{x_i x_i}(x) \approx \frac{f(x + e_i \Delta x_i) - 2f(x) + f(x - e_i \Delta x_i)}{\Delta x_i^2} = \frac{f_{i+} - 2f + f_{i-}}{\Delta x_i^2} \quad i = 1, 2, \ldots, n$$

$$f_{x_i x_j}(x) \approx \frac{f_{i+j+} - f_{i+j-} - f_{i-j+} + f_{i-j-}}{4 \Delta x_i \Delta x_j} \qquad i = j+1, j+2, \ldots, n; \; j = 1, 2, \ldots, n$$

$$f_{x_j x_i}(x) = f_{x_i x_j}(x) \qquad i = j+1, j+2, \ldots, n; \; j = 1, 2, \ldots, n$$

- Function evaluations: $1 + 2n + 4(n-1)n/2 = 1 + 2n + 2(n-1)n$
  Function: $f = f(x)$
  Gradient: $f_{i+} = f(x + e_i \Delta x_i)$, $f_{i-} = f(x - e_i \Delta x_i)$, i=1,2, ..., n
  Hessian: $f_{i+j+} = f(x + e_i \Delta x_i + e_j \Delta x_j)$, $f_{i+j-} = f(x + e_i \Delta x_i - e_j \Delta x_j)$,
  $f_{i-j+} = f(x - e_i \Delta x_i + e_j \Delta x_j)$, $f_{i-j-} = f(x - e_i \Delta x_i - e_j \Delta x_j)$

Calculating Derivatives

Finite Difference Approximation of the Gradient

$$f : \mathbb{R}^n \mapsto \mathbb{R} \qquad g(x) = \nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$$

Forward Difference Approximation

$$\frac{\partial f}{\partial x_i}(x) \approx g_{FD,i}(x) = \frac{f(x + h_i e_i) - f(x)}{h_i} \qquad h_i = \sqrt{u}\,(1 + |x_i|)$$

Backward Difference Approximation

$$\frac{\partial f}{\partial x_i}(x) \approx g_{BD,i}(x) = \frac{f(x) - f(x - h_i e_i)}{h_i} \qquad h_i = \sqrt{u}\,(1 + |x_i|)$$

Central Difference Approximation

$$\frac{\partial f}{\partial x_i}(x) \approx g_{CD,i}(x) = \frac{f(x + h_i e_i) - f(x - h_i e_i)}{2h_i} \qquad h_i = u^{1/3}\,(1 + |x_i|)$$

Machine Precision (double precision): $u = 2^{-52} \approx 2.22 \cdot 10^{-16}$

Matlab - Gradient - Forward Difference

$$\frac{\partial f}{\partial x_i}(x) \approx g_{FD,i}(x) = \frac{f(x + h_i e_i) - f(x)}{h_i} \qquad h_i = \sqrt{u}\max\{1.0, |x_i|\}$$

```
function g = gradientFD(fun,f,x,varargin)

pert = sqrt(eps);
nx   = length(x);

g = zeros(nx,1);
for i=1:nx
    h     = pert*max( 1.0, abs(x(i)) );
    xh    = x;
    xh(i) = xh(i) + h;
    h     = xh(i)-x(i);
    fh    = feval(fun,xh,varargin{:});
    g(i)  = (fh-f)/h;
end
```

Matlab - Gradient - Backward Difference

$$\frac{\partial f}{\partial x_i}(x) \approx g_{BD,i}(x) = \frac{f(x) - f(x - h_i e_i)}{h_i} \qquad h_i = \sqrt{u}\max\{1.0, |x_i|\}$$

```
function g = gradientBD(fun,f,x,varargin)

pert = sqrt(eps);
nx   = length(x);

g = zeros(nx,1);
for i=1:nx
    h     = pert*max( 1.0, abs(x(i)) );
    xh    = x;
    xh(i) = xh(i) - h;
    h     = x(i) - xh(i);
    fh    = feval(fun,xh,varargin{:});
    g(i)  = (f-fh)/h;
end
```

Matlab - Gradient - Central Difference

$$\frac{\partial f}{\partial x_i}(x) \approx g_{CD,i}(x) = \frac{f(x + h_i e_i) - f(x - h_i e_i)}{2h_i} \qquad h_i = u^{1/3} \max\{1.0, |x_i|\}$$

```matlab
function g = gradientCD(fun,f,x,varargin)

pert = (eps)^(1.0/3.0);
nx   = length(x);

g = zeros(nx,1);
for i=1:nx
    h      = pert*max( 1.0, abs(x(i)) );
    xph    = x;
    xph(i) = xph(i) + h;
    xmh    = x;
    xmh(i) = xmh(i) - h;
    dx     = xph(i) - xmh(i);
    fph    = feval(fun,xph,varargin{:});
    fmh    = feval(fun,xmh,varargin{:});
    g(i)   = (fph-fmh)/dx;
end
```

Finite Difference Approximation of the Jacobian Matrix

$$c : \mathbb{R}^n \mapsto \mathbb{R}^m \quad c(x) = \begin{bmatrix} c_1(x) \\ c_2(x) \\ \vdots \\ c_m(x) \end{bmatrix} \quad J(x) = \begin{bmatrix} \frac{\partial c_1}{\partial x_1}(x) & \frac{\partial c_1}{\partial x_2}(x) & \dots & \frac{\partial c_1}{\partial x_n}(x) \\ \frac{\partial c_2}{\partial x_1}(x) & \frac{\partial c_2}{\partial x_2}(x) & \dots & \frac{\partial c_2}{\partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial c_m}{\partial x_1}(x) & \frac{\partial c_m}{\partial x_2}(x) & \dots & \frac{\partial c_m}{\partial x_n}(x) \end{bmatrix}$$

Forward Difference Approximation

$$[J(x)]_{:,i} = \frac{\partial}{\partial x_i} c(x) \approx \frac{c(x + h_i e_i) - c(x)}{h_i} \qquad h_i = \sqrt{u}\,(1 + |x_i|)$$

Backward Difference Approximation

$$[J(x)]_{:,i} = \frac{\partial}{\partial x_i} c(x) \approx \frac{c(x) - c(x - h_i e_i)}{h_i} \qquad h_i = \sqrt{u}\,(1 + |x_i|)$$

Central Difference Approximation

$$[J(x)]_{:,i} = \frac{\partial}{\partial x_i} c(x) \approx \frac{c(x + h_i e_i) - c(x - h_i e_i)}{2h_i} \qquad h_i = u^{1/3}\,(1 + |x_i|)$$

Matlab - Jacobian - Forward Difference

$$[J(x)]_{:,i} = \frac{\partial}{\partial x_i} c(x) \approx \frac{c(x + h_i e_i) - c(x)}{h_i} \qquad h_i = \sqrt{u}\,\max\{1.0, |x_i|\}$$

```
function J = JacobianFD(cfun,c,x,varargin)

pert = sqrt(eps);
nx   = length(x);
nc   = length(c);

J = zeros(nc,nx);
for i=1:nx
    h       = pert*max( 1.0, abs(x(i)) );
    xh      = x;
    xh(i)   = xh(i) + h;
    h       = xh(i)-x(i);
    ch      = feval(cfun,xh,varargin{:});
    J(:,i)  = (ch-c)/h;
end
```

Matlab - Jacobian - Backward Difference

$$[J(x)]_{:,i} = \frac{\partial}{\partial x_i} c(x) \approx \frac{c(x) - c(x - h_i e_i)}{h_i} \qquad h_i = \sqrt{u}\max\{1.0, |x_i|\}$$

```
function J = JacobianBD(cfun,c,x,varargin)

pert = sqrt(eps);
nx   = length(x);
nc   = length(c);

J = zeros(nc,nx);
for i=1:nx
    h      = pert*max( 1.0, abs(x(i)) );
    xh     = x;
    xh(i)  = xh(i) - h;
    h      = x(i)-xh(i);
    ch     = feval(cfun,xh,varargin{:});
    J(:,i) = (c-ch)/h;
end
```

Matlab - Jacobian - Central Difference

$$[J(x)]_{:,i} = \frac{\partial}{\partial x_i} c(x) \approx \frac{c(x + h_i e_i) - c(x - h_i e_i)}{2h_i} \qquad h_i = u^{1/3} \max\{1.0, |x_i|\}$$

```
function J = JacobianCD(cfun,c,x,varargin)

pert = (eps)^(1.0/3.0);
nx   = length(x);
nc   = length(c);

J = zeros(nc,nx);
for i=1:nx
    h      = pert*max( 1.0, abs(x(i)) );
    xph    = x;
    xph(i) = xph(i) + h;
    xmh    = x;
    xmh(i) = xmh(i) - h;
    dx     = xph(i) - xmh(i);
    cph    = feval(cfun,xph,varargin{:});
    cmh    = feval(cfun,xmh,varargin{:});
    J(:,i) = (cph-cmh)/dx;
end
```

The Hessian Matrix

- Vector: $x \in \mathbb{R}^n$
- Objective function: $f : \mathbb{R}^n \mapsto \mathbb{R}$

$$f = f(x)$$

- Gradient: $g : \mathbb{R}^n \mapsto \mathbb{R}^n$

$$g = g(x) = \nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$$

- Hessian matrix: $H : \mathbb{R}^n \mapsto \mathbb{R}^{n \times n}$

$$H = H(x) = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1}\right)(x) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1}\right)(x) & \dots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_1}\right)(x) \\ \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right)(x) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_2}\right)(x) & \dots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_2}\right)(x) \\ \\ \vdots \\ \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_n}\right)(x) & \frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_n}\right)(x) & \dots & \frac{\partial}{\partial x_n}\left(\frac{\partial f}{\partial x_n}\right)(x) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1}\nabla f(x) & \frac{\partial}{\partial x_2}\nabla f(x) & \dots & \frac{\partial}{\partial x_n}\nabla f(x) \end{bmatrix}$$

Finite Difference Approximation of the Hessian Matrix

- Gradient evaluations, $g = g(x) = \nabla f(x)$, available

$$f : \mathbb{R}^n \mapsto \mathbb{R}, \quad f = f(x)$$

$$g = g(x) = \nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$$

$$H = H(x) = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1}\nabla f(x) & \frac{\partial}{\partial x_2}\nabla f(x) & \dots & \frac{\partial}{\partial x_n}\nabla f(x) \end{bmatrix}$$

- Forward Difference Approximation: `H = JacobianFD(@gfun,g,x)`

$$[H(x)]_{:,i} = \frac{\partial}{\partial x_i}\nabla f(x) \approx \frac{\nabla f(x + h_i e_i) - \nabla f(x)}{h_i} \qquad h_i = \sqrt{u}\,(1 + |x_i|)$$

- Backward Difference Approximation: `H = JacobianBD(@gfun,g,x)`

$$[H(x)]_{:,i} = \frac{\partial}{\partial x_i}\nabla f(x) \approx \frac{\nabla f(x) - \nabla f(x - h_i e_i)}{h_i} \qquad h_i = \sqrt{u}\,(1 + |x_i|)$$

- Central Difference Approximation: `H = JacobianCD(@gfun,g,x)`

$$[H(x)]_{:,i} = \frac{\partial}{\partial x_i}\nabla f(x) \approx \frac{\nabla f(x + h_i e_i) - \nabla(x - h_i e_i)}{2h_i} \qquad h_i = u^{1/3}\,(1 + |x_i|)$$

- Hessian symmetric

$$H := \frac{H + H'}{2}$$

Finite Difference Approximation of the Hessian Matrix

Only function evaluations, $f(x)$, available

$$f : \mathbb{R}^n \mapsto \mathbb{R} \quad \nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix} \quad \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{bmatrix}$$

$$A_{ij}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \approx \frac{f(x + h_i e_i + h_j e_j) - f(x + h_i e_i) - f(x + h_j e_j) + f(x)}{h_i h_j}$$

$$h_i = \sqrt{u}(1 + |x_i|)$$
$$h_j = \sqrt{u}(1 + |x_j|)$$

$$\nabla^2 f(x) \approx \frac{A(x) + A(x)^T}{2}$$

Alternatively, one can evaluate the lower triangular part of $\nabla^2 f(x)$ only.