Unconstrained Problem and Algorithm III

Jong-June Jeon

October 13, 2025

Department of Statistics, University of Seoul

Hessian matrix approximation

Let $f : \mathbb{R}^2 \mapsto \mathbb{R}^2$ and denote the image of f by $(f_1(x_1, x_2), f_2(x_1, x_2))$, where $f_j : \mathbb{R}^2 \mapsto \mathbb{R}$

ex) $f(x_1, x_2) = (x_1, x_1^2 + x_2)$

How to write the change of output f according to a small perturbation? $\rightarrow$ Jacobian!

$$J_f(x) = \begin{pmatrix} \frac{\partial f_1(x_1,x_2)}{\partial x_1} & \frac{\partial f_1(x_1,x_2)}{\partial x_2} \\ \frac{\partial f_2(x_1,x_2)}{\partial x_1} & \frac{\partial f_2(x_1,x_2)}{\partial x_2} \end{pmatrix}$$

Composition and derivatives

Let $h : \mathbb{R}^p \mapsto \mathbb{R}^q$ and $f : \mathbb{R}^q \mapsto \mathbb{R}$ (f and h are continuously differentiable.)

$$\frac{\partial}{\partial x_1} f(h(x_1, \cdots, x_p))?$$

Let $h(x_1, \cdots, x_p) = (h_1(x_1, \cdots, x_p), \cdots, h_p(x_1, \cdots, x_p))$ where $h_j : \mathbb{R}^p \mapsto \mathbb{R}$. Jacobian of h is given by

$$J_h(x) = \begin{pmatrix} \frac{\partial h_1(x)}{\partial x_1} & \dots & \frac{\partial h_1(x)}{\partial x_p} \\ \vdots & \vdots & \vdots \\ \frac{\partial h_q(x)}{\partial x_1} & \dots & \frac{\partial h_q(x)}{\partial x_p} \end{pmatrix}$$

Composition and derivatives

Let $z = f(u_1, \cdots, u_q)$. By fundamental lemma

$$dz = \frac{\partial z}{u_1}\Delta u_1 + \cdots + \frac{\partial z}{u_q}\Delta u_q.$$

Let $u_j = h_j(x)$ then

$$\begin{aligned}\frac{\partial}{\partial x_1} f(h(x_1, \cdots, x_p)) &= \sum_{j=1}^{q} \frac{\partial f(u)}{\partial u_j}\frac{\partial h_j(x)}{\partial x_1} \\ &= (J_h(x)^\top \nabla f(u))[0, :]\end{aligned}$$

Composition and derivatives

$$\frac{\partial f(h)}{\partial x} = J_h(x)^\top \nabla f(u)$$

Jacobian and Hessian

Let $f : x \in \mathbb{R}^p \mapsto \mathbb{R}$ and $\nabla f = (f_1, \cdots, f_p)^\top$, where f_i is the derivative of f. The Hessian matrix is the Jacobian matrix of ∇f.

$$J_{\nabla f}(x) = \frac{\nabla f}{\partial x^\top} = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \dots & \frac{\partial f_1(x)}{\partial x_p} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_p(x)}{\partial x_1} & \dots & \frac{\partial f_p(x)}{\partial x_p} \end{pmatrix} = \nabla^2 f$$

The Hessian matrix of f represents the change of ∇f according to a small perturbation of x.

Directional derivatives

Let $f : \mathbb{R}^p \mapsto \mathbb{R}$ and let $v \in \mathbb{R}^p$. Denote $f_j(x) = \frac{\partial f(x)}{\partial x_j}$.

$$\frac{\partial f(x+tv)}{\partial t} = \sum_{j=1}^{p} f_j(x+tv)v_j = \nabla f(x+tv)^\top v$$

Thus, the direction derivatives along to v are equal to $\nabla f(x)^\top v$.

Directional derivatives

Conversely $\nabla f(x)^\top v$ is approximated by the direction derivatives along to v:

$$\nabla f(x)^\top v \simeq \frac{f(x+tv) - f(x)}{t}$$

Hessian and Directional derivatives

$$(\nabla^2 f)v = J_{\nabla f}(x)v = \begin{pmatrix} \sum_{j=1}^{p} \frac{\partial f_1(x)}{\partial x_j} v_j \\ \vdots \\ \sum_{j=1}^{p} \frac{\partial f_p(x)}{\partial x_j} v_j \end{pmatrix} = \begin{pmatrix} \nabla f_1(x)^\top v \\ \vdots \\ \nabla f_p(x)^\top v \end{pmatrix} \simeq \begin{pmatrix} (f_1(x+tv) - f_1(x))/t \\ \vdots \\ (f_p(x+tv) - f_1(x))/t \end{pmatrix}$$

as $t \to 0$. In summary,

$$(\nabla^2 f)v \simeq t^{-1}(\nabla f(x+tv) - \nabla f(x))$$

as $t \to 0$.

Sherman–Morrison Formula

Sherman–Morrison Formula Statement. Let $A \in \mathbb{R}^{n\times n}$ be invertible and $u, v \in \mathbb{R}^n$. If $1 + v^\top A^{-1} u \neq 0$, then

$$(A + uv^\top)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}.$$

Condition:

$$1 + v^\top A^{-1} u \neq 0 \quad \text{(otherwise } A + uv^\top \text{ is singular).}$$

Interpretation:

- Efficiently updates the inverse when A is modified by a rank-1 matrix $uv^\top$.
- Avoids recomputing the full inverse ($O(n^3)$); only needs $O(n^2)$ work.

Sketch of Derivation We seek $X = (A + uv^\top)^{-1}$ such that

$$(A + uv^\top)X = I.$$

Rearranging:

$$AX + u(v^\top X) = I \quad \Rightarrow \quad X = A^{-1} - A^{-1}u(v^\top X).$$

Multiply by $v^\top$:

$$v^\top X = v^\top A^{-1} - v^\top A^{-1}u(v^\top X) \Rightarrow (1 + v^\top A^{-1}u)v^\top X = v^\top A^{-1}.$$

Substitute back:

$$\boxed{X = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}}.$$

Extension: Sherman–Morrison–Woodbury Identity

For rank-k updates with $U, V \in \mathbb{R}^{n\times k}$ and $C \in \mathbb{R}^{k\times k}$:

$$(A + UCV^\top)^{-1} = A^{-1} - A^{-1}U(C^{-1} + V^\top A^{-1}U)^{-1}V^\top A^{-1}.$$

Special case: $k = 1$ and $C = 1 \Rightarrow$ original Sherman–Morrison formula.

Strong Wolfe Conditions

Strong Wolfe Conditions

Let $\phi(\alpha) = f(x_k + \alpha p_k)$ with descent direction p_k s.t. $\frac{\partial}{\partial \alpha} f(x_k + \alpha p_k)|_{\alpha=0} = \nabla f(x_k)^\top p_k < 0$.

Find a stepsize $\alpha_k > 0$ satisfying:

$$\text{(Armijo)} \quad \phi(\alpha_k) \leq \phi(0) + c_1\, \alpha_k\, \phi'(0),$$

$$\text{(Strong Wolfe curvature)} \quad \left|\phi'(\alpha_k)\right| \leq c_2 \left|\phi'(0)\right|.$$

Typical choices: $c_1 = 10^{-4}$, $c_2 \in [0.1, 0.9]$.

Role of Strong Wolfe in BFGS: Ensuring $y_k^\top s_k > 0$ (See Unconstrained Problem and Algorithm III).

Let $p_k = x_{k+1} - x_x$ and assume that $\nabla f(x_k)^\top p_k < 0$. Define $\phi(\alpha) = f(x_k + \alpha p_k)$, so $\phi'(0) = \nabla f(x_k)^\top p_k < 0$ and

$$y_k^\top s_k = \alpha_k \left(\phi'(\alpha_k) - \phi'(0)\right).$$

Strong Wolfe curvature gives $|\phi'(\alpha_k)| \leq c_2 |\phi'(0)|$ with $0 < c_2 < 1$, hence $\phi'(\alpha_k) \geq c_2\, \phi'(0)$. Therefore,

$$y_k^\top s_k = \alpha_k(\phi'(\alpha_k) - \phi'(0)) \geq \alpha_k(c_2 - 1)\phi'(0) > 0.$$

Conclusion: Strong Wolfe + descent direction $\Rightarrow y_k^\top s_k > 0$.

2) If $s_k^\top y_k > 0$, then BFGS preserves positive definiteness (PD) BFGS update in the inverse-Hessian form $(B_k \succ 0)$:

$$\rho_k := \frac{1}{y_k^\top s_k} > 0, \qquad B_{k+1} = (I - \rho_k s_k y_k^\top) B_k (I - \rho_k y_k s_k^\top) + \rho_k s_k s_k^\top.$$

For any nonzero vector $z \neq 0$:

$$z^\top B_{k+1} z = \underbrace{\big((I - \rho_k y_k s_k^\top)z\big)^\top B_k \big((I - \rho_k y_k s_k^\top)z\big)}_{\geq 0} + \rho_k (s_k^\top z)^2 \ (\geq 0).$$

- The first term is > 0 since $B_k \succ 0$ and $(I - \rho_k y_k s_k^\top)z \neq 0$.
- If $(I - \rho_k y_k s_k^\top)z = 0$, then $z = \rho_k y_k (s_k^\top z)$. In this case, $z \neq 0 \Rightarrow s_k^\top z \neq 0$, so the second term $\rho_k (s_k^\top z)^2 > 0$.

Hence, both terms cannot vanish simultaneously:

$$z^\top B_{k+1} z > 0 \ \ \forall z \neq 0 \ \Rightarrow \ B_{k+1} \succ 0.$$