



بهینه سازی
مبانی بهینه سازی نامقید
روش های نیوتن و شبه نیوتن

محسن هوشمند
دانشکده تکنولوژی اطلاعات و علم رایانه
دانشگاه تحصیلات تکمیلی علوم پایه زنجان

روش‌های تاکنون

گرادیان نزولی مبتنی بر مشتق مرتبه اول

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{x}_i - \alpha_i (\nabla f(\mathbf{x}_i)) \Leftarrow \text{گن} \\ \mathbf{x}_{i+1} &= \mathbf{x}_i + \alpha_i (-\nabla f(\mathbf{x}_i) + \beta_i \mathbf{p}_i) \Leftarrow \text{گم} \end{aligned}$$

روشی مبتنی بر مشتق مرتبه دوم

▪ ماتریس هسی

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha_i H^{-1} \nabla f(\mathbf{x}_i) \Leftarrow \text{نیوتن}$$

روش‌های تاکنون

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

روش‌های تاکنون

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

تاکنون

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha_i V_i \nabla f(\mathbf{x}_i)^T$$

$$V_i = I$$

؟ ▪

$$V_i = H^{-1}$$

؟ ▪

روش نیوتن

$$f(\mathbf{x}_i + \mathbf{p}_i) \approx f(\mathbf{x}_i) + \nabla f(\mathbf{x}_i)^T \mathbf{p}_i + \frac{1}{2} \mathbf{p}_i^T \nabla^2 f(\mathbf{x}_i) \mathbf{p}_i$$
$$f(\mathbf{x}_i + \mathbf{p}_i) \approx \underbrace{f(\mathbf{x}_i) + \mathbf{g}_i^T \mathbf{p}_i + \frac{1}{2} \mathbf{p}_i^T \mathbf{H}_i \mathbf{p}_i}_{q(\mathbf{p})}$$

روش نیوتن: \mathbf{p} کاهنده تقریب درجه دو $q(\mathbf{p})$

$$\nabla q(\mathbf{p}_i) = \mathbf{g} + \mathbf{H} \mathbf{p}_i = \mathbf{0}$$

معادله نیوتن

$$\mathbf{H} \mathbf{p} = -\mathbf{g}$$
$$\mathbf{p}_i = -\mathbf{H}_i^{-1} \mathbf{g}_i$$

\mathbf{p}_i جهت نیوتن

روش شبه نیوتن

- و دیویدن
- آزمایشگاه ملی آرگونه
- در دهه ۳۰ شمسی
- اولین شماره سیام درباره بهینه‌سازی در سال ۱۳۷۰ شمسی

بین روش تندترین نزول و روش نیوتن
غیرعملی بودن محاسبه ماتریس هسی
تقریب ماتریس وارون هسی

روش شبه نیوتن - الگوریتم

مرحله i -ام

$$\mathbf{p}_i = -B_i^{-1} \mathbf{g}_i$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i \text{ با جستجو خط (ناکامل)}$$

$$\mathbf{g}_{i+1} = \nabla f(\mathbf{x}_{i+1})$$

$$B_{i+1} \propto (B_i, \mathbf{x}_{i+1} - \mathbf{x}_i, \mathbf{g}_{i+1} - \mathbf{g}_i)$$

روش شبه نیوتن

ماتریس هسی

- هزینه بالا یا غیرممکنی محاسبه دقیق هسی

چگونه؟

- تقریبی هسی
- B_i دارای اطلاع
- استفاده از نو کردن مناسب
- بروز کردن ماتریس در هر تکرار با استفاده از اطلاع مرتبه اول
- استفاده از منحنی اندازه گیری شده در طول قدم فعلی

روش شبه نیوتن

$$m_i(\mathbf{p}) = f_i + \mathbf{p}^T \nabla f_i + \mathbf{p}^T B_i \mathbf{p}$$

- B_i تخمینی از $\nabla^2 f(\mathbf{x}_i)$
- ماتریس مثبت معین متقارن
- بروز شدن مقدار ماتریس در هر تکرار

$$\nabla m_i(\mathbf{p}) = \nabla f_i + B_i \mathbf{p}$$

$$\mathbf{p} = \mathbf{0}$$

$$m_i(\mathbf{0}) = f_i$$

$$\nabla m_i(\mathbf{0}) = \nabla f_i$$

$m_i(\mathbf{p})$ تابع درجه دو کوژ

- کمینه ساز تابع

$$\mathbf{p}_i = -B_i^{-1} \nabla f_i$$

▪ $\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$ با ضریبی برآورده کننده شرطهای وولف

- مشابه روش نیوتن

با تخمین به جای مقدار معکوس هسی

روش شبه نیوتن

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

$$\mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

ایجاد مدل جدید حول نقطه \mathbf{x}_{i+1} استفاده از

$$m_{i+1}(\mathbf{p}) = f_{i+1} + \mathbf{p}^T \nabla f_{i+1} + \mathbf{p}^T B_{i+1} \mathbf{p}$$

به دنبال B_{i+1}

▪ بر اساس دانش بدست آمده در مراحل آخر

▪ برابری گرادیان m_{i+1} در دو نقطه \mathbf{x}_{i+1} و \mathbf{x}_i

$$\nabla m_{i+1}(\mathbf{0}) = \nabla f_{i+1} \quad \square$$

$$\nabla m_{i+1}(-\alpha_i \mathbf{p}_i) = \nabla f_{i+1} - B_{i+1} \mathbf{s}_i \quad \square$$

$$\nabla m_{i+1}(-\alpha_i \mathbf{p}_i) = \nabla f_i \quad \square$$

$$\nabla f_i = \nabla f_{i+1} - B_{i+1} \mathbf{s}_i \quad \square$$

$$B_{i+1} \mathbf{s}_i = \nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i)$$

▪ معروف به «معادله وتر»

روش شبه نیوتن - ادامه

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

$$\mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{g}_i = \nabla f(\mathbf{x}_i)$$

$$\mathbf{y}_i = \mathbf{g}_{i+1} - \mathbf{g}_i$$

چگونه؟

▪ B_i دارای اطلاع

▪ نو کردن مناسب با استفاده از اطلاعات قبلی

محاسبه B_{i+1}

$$\nabla f(\mathbf{x}_{i+1}) = \nabla f(\mathbf{x}_i) - B_{i+1} \mathbf{s}_i$$

یا

$$B_{i+1} \mathbf{s}_i = \nabla f(\mathbf{x}_{i+1}) - \nabla f(\mathbf{x}_i)$$

معروف به «معادله وتر» (نمایش‌های دیگر)

$$B_{i+1} \mathbf{s}_i = \mathbf{g}_{i+1} - \mathbf{g}_i$$

$$B_{i+1} \mathbf{s}_i = \mathbf{y}_i$$

روش شبه نیوتن - نو کردن متقارن رتبه-یک

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

$$\mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{g}_i = \nabla f(\mathbf{x}_i)$$

$$\mathbf{y}_i = \mathbf{g}_{i+1} - \mathbf{g}_i$$

$$B_{i+1} \mathbf{s}_i = \mathbf{y}_i$$

$$B_{i+1} = B_i + a \mathbf{u} \mathbf{u}^T$$

نو کردن با ماتریس رتبه یک

$$a = -1 \text{ یا } a = +1 \quad \blacksquare$$

انتخاب a و \mathbf{u} به نحوی که برآورده ساز معادله وتر باشد

$$\begin{aligned} B_{i+1} \mathbf{s}_i = \mathbf{y}_i &\Rightarrow B_i \mathbf{s}_i + a \mathbf{u} \mathbf{u}^T \mathbf{s}_i = \mathbf{y}_i \\ &\Rightarrow (a \mathbf{u}^T \mathbf{s}_i) \mathbf{u} = \mathbf{y}_i - B_i \mathbf{s}_i \end{aligned}$$

$\mathbf{u}^T \mathbf{s}_i$ صرفاً ضریب، پس جابجاپذیر

پس \mathbf{u} ضریبی از $\mathbf{y}_i - B_i \mathbf{s}_i$

$$\mathbf{u} = \zeta (\mathbf{y}_i - B_i \mathbf{s}_i) \quad \blacksquare$$

$$\begin{aligned} \mathbf{y}_i - B_i \mathbf{s}_i &= a \zeta^2 (\mathbf{y}_i - B_i \mathbf{s}_i) (\mathbf{y}_i - B_i \mathbf{s}_i)^T \mathbf{s}_i \\ \mathbf{y}_i - B_i \mathbf{s}_i &= a \zeta^2 [\mathbf{s}_i^T (\mathbf{y}_i - B_i \mathbf{s}_i)] (\mathbf{y}_i - B_i \mathbf{s}_i) \end{aligned}$$

روش شبه نیوتن - نو کردن متقارن رتبه-یک

$$\begin{aligned} \mathbf{y}_i - B_i \mathbf{s}_i &= a \zeta^2 (\mathbf{y}_i - B_i \mathbf{s}_i) (\mathbf{y}_i - B_i \mathbf{s}_i)^T \mathbf{s}_i \\ \mathbf{y}_i - B_i \mathbf{s}_i &= a \zeta^2 [\mathbf{s}_i^T (\mathbf{y}_i - B_i \mathbf{s}_i)] (\mathbf{y}_i - B_i \mathbf{s}_i) \end{aligned}$$

$$\begin{aligned} a & \cdot \\ [\mathbf{s}_i^T (\mathbf{y}_i - B_i \mathbf{s}_i)] & \text{ برابر علامت} \cdot \\ \zeta & \cdot \\ \pm [\mathbf{s}_i^T (\mathbf{y}_i - B_i \mathbf{s}_i)]^{-\frac{1}{2}} & \cdot \end{aligned}$$

نو کردن با ماتریس رتبه یک

$$B_{i+1} = B_i + a \mathbf{u} \mathbf{u}^T$$

$$\mathbf{u} = \zeta (\mathbf{y}_i - B_i \mathbf{s}_i)$$

نتیجه: نو کردن متقارن رتبه-یک برآورده کننده معادله وتر

$$B_{i+1} = B_i + \frac{(\mathbf{y}_i - B_i \mathbf{s}_i) (\mathbf{y}_i - B_i \mathbf{s}_i)^T}{(\mathbf{y}_i - B_i \mathbf{s}_i)^T \mathbf{s}_i}$$

نوکردن متقارن رتبه-یک (مر-۱) - ادامه

در واقع به دنبال معکوس B_{i+1}

قضیه معادله شرم-موريسن-وودباری

اگر $A \in \mathbb{R}^{n \times n}$ و $u, v \in \mathbb{R}^n$ و $A + uv^T$ ناتکین باشد، آن گاه معکوس آن برابر است با

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

اثبات-

نوکردن متقارن رتبه-یک (مر-۱) - ادامه

در واقع به دنبال معکوس B_{i+1}

تعمیم قضیه معادله شرمن-موریسن-وودباری

اگر $A \in \mathbb{R}^{n \times n}$ و $U, V \in \mathbb{R}^{n \times d}$ آن گاه $A + UV^T$ معکوس پذیر خواهد بود اگر و فقط اگر $I + V^T A^{-1} U$ معکوس پذیر باشد. در این صورت:

$$(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + V^T A^{-1}U)^{-1}V^T A^{-1}$$

نوکردن متقارن رتبه-یک (مر-۱) - ادامه

$$B_{i+1} = B_i + \frac{(\mathbf{y}_i - B_i \mathbf{s}_i)(\mathbf{y}_i - B_i \mathbf{s}_i)^T}{(\mathbf{y}_i - B_i \mathbf{s}_i)^T \mathbf{s}_i}$$

در واقع به دنبال معکوس B_{i+1}

قضیه معادله شرم-موریسن-وودباری

اگر $A \in \mathbb{R}^{n \times n}$ و $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ و $A + \mathbf{u}\mathbf{v}^T$ ناتکین باشد، آن گاه معکوس آن برابر است با

$$(A + \mathbf{u}\mathbf{v}^T)^{-1} = A^{-1} - \frac{A^{-1} \mathbf{u}\mathbf{v}^T A^{-1}}{1 + \mathbf{v}^T A^{-1} \mathbf{u}}$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

$$\mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{g}_i = \nabla f(\mathbf{x}_i)$$

$$\mathbf{y}_i = \mathbf{g}_{i+1} - \mathbf{g}_i$$

$$B_{i+1} \mathbf{s}_i = \mathbf{y}_i$$

آن گاه مر-۱ B^{\wedge} یا معکوس B :

$$B_{i+1}^{\wedge} = B_i^{\wedge} + \frac{(\mathbf{s}_i - B_i^{\wedge} \mathbf{y}_i)(\mathbf{s}_i - B_i^{\wedge} \mathbf{y}_i)^T}{(\mathbf{s}_i - B_i^{\wedge} \mathbf{y}_i)^T \mathbf{y}_i}$$

نوکردن متقارن رتبه-یک (مر-۱) - ادامه

$$B_{i+1} = B_i + \frac{(\mathbf{y}_i - B_i \mathbf{s}_i)(\mathbf{y}_i - B_i \mathbf{s}_i)^T}{(\mathbf{y}_i - B_i \mathbf{s}_i)^T \mathbf{s}_i}$$

آن گاه مر-۱ $B^{\hat{}}$ یا معکوس B :

$$B_{i+1}^{\hat{}} = B_i^{\hat{}} + \frac{(\mathbf{s}_i - B_i^{\hat{}} \mathbf{y}_i)(\mathbf{s}_i - B_i^{\hat{}} \mathbf{y}_i)^T}{(\mathbf{s}_i - B_i^{\hat{}} \mathbf{y}_i)^T \mathbf{y}_i}$$

نوکردن متقارن رتبه-یک (مر-۱) - ادامه

امکان صفر بودن $(\mathbf{y}_i - B_i \mathbf{s}_i)^T \mathbf{s}_i$

عدم رعایت مثبت معین بودن

$$B_{i+1}^{\hat{}} = B_i^{\hat{}} + \frac{(\mathbf{s}_i - B_i^{\hat{}} \mathbf{y}_i)(\mathbf{s}_i - B_i^{\hat{}} \mathbf{y}_i)^T}{(\mathbf{s}_i - B_i^{\hat{}} \mathbf{y}_i)^T \mathbf{y}_i}$$

روش‌های نوکردن تکراری شبه‌نیوتن دیگر

۱-۴

$$B_{i+1}^{\hat{r}} = B_i^{\hat{r}} + \frac{(\mathbf{s}_i - B_i^{\hat{r}} \mathbf{y}_i)(\mathbf{s}_i - B_i^{\hat{r}} \mathbf{y}_i)^T}{(\mathbf{s}_i - B_i^{\hat{r}} \mathbf{y}_i)^T \mathbf{y}_i}$$

روش‌های نوکردن تکراری شبه‌نیوتن دیگر

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

$$\mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{g}_i = \nabla f(\mathbf{x}_i)$$

$$\mathbf{y}_i = \mathbf{g}_{i+1} - \mathbf{g}_i$$

$$B_{i+1} \mathbf{s}_i = \mathbf{y}_i$$

دیویدن-فلچر-پاول DFP

$$B_{i+1}^{\hat{r}} = B_i^{\hat{r}} - \frac{B_i^{\hat{r}} \mathbf{y}_i \mathbf{y}_i^T B_i^{\hat{r}}}{\mathbf{y}_i^T B_i^{\hat{r}} \mathbf{y}_i} + \frac{\mathbf{s}_i \mathbf{s}_i^T}{\mathbf{y}_i^T \mathbf{s}_i}$$

روش‌های نو کردن تکراری شبه‌نیوتن دیگر

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

$$\mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{g}_i = \nabla f(\mathbf{x}_i)$$

$$\mathbf{y}_i = \mathbf{g}_{i+1} - \mathbf{g}_i$$

$$B_{i+1} \mathbf{s}_i = \mathbf{y}_i$$

برویدن-فلچر-گولدفرب-شانو BFGS

$$\gamma_i = \frac{1}{\mathbf{y}_i^T \mathbf{s}_i}$$

$$B_{i+1}^{\hat{}} = (I - \gamma_i \mathbf{s}_i \mathbf{y}_i^T) B_i^{\hat{}} (I - \gamma_i \mathbf{y}_i \mathbf{s}_i^T) + \gamma_i \mathbf{s}_i \mathbf{s}_i^T$$

\mathbf{s}_i یا \mathbf{p}_i

روش‌های جستجو خط

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{x}_i + \alpha_i \mathbf{p}_i \\ \mathbf{x}_{i+1} - \mathbf{x}_i &= \alpha_i \mathbf{p}_i \\ \mathbf{s}_i &= \mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i \\ \mathbf{g}_i &= \nabla f(\mathbf{x}_i) \\ \mathbf{y}_i &= \mathbf{g}_{i+1} - \mathbf{g}_i \\ B_{i+1} \mathbf{s}_i &= \mathbf{y}_i \\ \gamma_i &= \frac{1}{\mathbf{y}_i^T \mathbf{s}_i} \end{aligned}$$

م-۱

$$B_{i+1}^{\hat{}} = B_i^{\hat{}} + \frac{(\mathbf{s}_i - B_i^{\hat{}} \mathbf{y}_i)(\mathbf{s}_i - B_i^{\hat{}} \mathbf{y}_i)^T}{(\mathbf{s}_i - B_i^{\hat{}} \mathbf{y}_i)^T \mathbf{y}_i}$$

د ف پ

$$B_{i+1}^{\hat{}} = B_i^{\hat{}} - \frac{B_i^{\hat{}} \mathbf{y}_i \mathbf{y}_i^T B_i^{\hat{}}}{\mathbf{y}_i^T B_i^{\hat{}} \mathbf{y}_i} + \frac{\mathbf{s}_i \mathbf{s}_i^T}{\mathbf{y}_i^T \mathbf{s}_i}$$

ب

$$B_{i+1}^{\hat{}} = B_i^{\hat{}} + \frac{(\mathbf{s}_i - B_i^{\hat{}} \mathbf{y}_i) \mathbf{s}_i^T B_i^{\hat{}}}{\mathbf{s}_i^T B_i^{\hat{}} \mathbf{y}_i}$$

ب ف گ ش

$$B_{i+1}^{\hat{}} = (I - \gamma_i \mathbf{s}_i \mathbf{y}_i^T) B_i^{\hat{}} (I - \gamma_i \mathbf{y}_i \mathbf{s}_i^T) + \gamma_i \mathbf{s}_i \mathbf{s}_i^T$$

الگوریتم

مقداردهی اولیه

$$B_0^f = I \quad \bullet$$

$$B_0^f = H^{-1} \quad \bullet$$

اگر تابع درجه دو باشد

▪ ماتریس هسی دقیق بعد از n مرحله به دست خواهد آمد

▪ متناظر با گم زمانی که $B_0^f = I$

اگر تابع غیردرجه دو

▪ شن معمولاً سریعتر از گم

▪ عدم نیاز به جستجو خط کامل

الگوریتم

انتخاب ϵ و \mathbf{x}_0

محاسبه ماتریس وارون هسی B_0^f

$i = 0$

تازمان $\|\nabla f_i\| > \epsilon$
▪ محاسبه جهت جستجو

$$\mathbf{p}_i = -B_i^f \nabla f_i$$
$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

$$\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$$
$$\mathbf{y}_i = \nabla f_{i+1} - \nabla f_i$$

$$\gamma_i = \frac{1}{\mathbf{y}_i^T \mathbf{s}_i}$$

$$B_{i+1}^f = (I - \gamma_i \mathbf{s}_i \mathbf{y}_i^T) B_i^f (I - \gamma_i \mathbf{y}_i \mathbf{s}_i^T) + \gamma_i \mathbf{s}_i \mathbf{s}_i^T$$

$$i = i + 1$$

یافتن α_i با جستجوی خط تا یافتن مقدار برآورده کننده شرط وولف

خانواده بریدون

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \alpha_i \mathbf{p}_i$$

$$\mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i = \alpha_i \mathbf{p}_i$$

$$\mathbf{g}_i = \nabla f(\mathbf{x}_i)$$

$$\mathbf{y}_i = \mathbf{g}_{i+1} - \mathbf{g}_i$$

$$B_{i+1} \mathbf{s}_i = \mathbf{y}_i$$

$$B_{i+1} = B_i - \frac{B_i \mathbf{s}_i \mathbf{s}_i^T B_i}{\mathbf{s}_i^T B_i \mathbf{s}_i} + \frac{\mathbf{y}_i \mathbf{y}_i^T}{\mathbf{y}_i^T \mathbf{s}_i} + \phi_i (\mathbf{s}_i^T B_i \mathbf{s}_i) \mathbf{v}_i \mathbf{v}_i^T$$

$$\mathbf{v}_i = \frac{\mathbf{y}_i}{\mathbf{y}_i^T \mathbf{s}_i} - \frac{B_i \mathbf{s}_i}{\mathbf{s}_i^T B_i \mathbf{s}_i}$$

برویدن-فلچر-گولدفرب-شانو

دیویدن-فلچر-پاول DFP

خانواده بریدون

$$B_{i+1} = B_i - \frac{B_i \mathbf{s}_i \mathbf{s}_i^T B_i}{\mathbf{s}_i^T B_i \mathbf{s}_i} + \frac{\mathbf{y}_i \mathbf{y}_i^T}{\mathbf{y}_i^T \mathbf{s}_i} + \phi_i (\mathbf{s}_i^T B_i \mathbf{s}_i) \mathbf{v}_i \mathbf{v}_i^T$$

$$\mathbf{v}_i = \frac{\mathbf{y}_i}{\mathbf{y}_i^T \mathbf{s}_i} - \frac{B_i \mathbf{s}_i}{\mathbf{s}_i^T B_i \mathbf{s}_i}$$

برویدن-فلچر-گولدفرب-شانو

$$\phi_i = 0 \quad \blacksquare$$

دیویدن-فلچر-پاول DFP

$$\phi_i = 1 \quad \blacksquare$$

خانواده بریدون

ترکیب

$$B_{i+1} = (1 - \phi_i)B_{i+1}^{BFGS} + \phi_i B_{i+1}^{DFP}$$

برویدن-فلچر-گولدفرب-شانو

$$\phi_i = 0 \quad \blacksquare$$

داویدون-فلچر-پاول DFP

$$\phi_i = 1 \quad \blacksquare$$

شبه نیوتن

هزینه بالا یا غیرممکنی محاسبه دقیق هسی

تقریبی هسی

بروز کردن ماتریس در هر تکرار با استفاده از اطلاع مرتبه اول

مزایا

- صرفا استفاده از مشتق مرتبه اول
- ماتریس تقریب وارون هسی مثبت معین
- نشانگر نزول
- $O(n^2)$ ضرب به ازای هر مرحله

معایب

- امکان نبود جهت نزول
- امکان دقیق نبودن تخمین هسی

منابع

[نازهدل]

[لوئین برگر]

[فلچر]

What are the differences between the different gradient-based numerical optimization methods?, <https://scicomp.stackexchange.com/questions/26960/what-are-the-differences-between-the-different-gradient-based-numerical-optimiza>

W. Hager, H. Zhang, "A Survey Of Nonlinear Conjugate Gradient Methods," Pac. J. Optim. 2, No. 1, 35-58, 2006.

"Conjugate Gradient," <https://bookdown.org/rdpeng/advstatcomp/conjugate-gradient.html>

M. Zibulevsky, "Method of Conjugate Gradients," https://www.youtube.com/watch?v=hZVK_PGE0_I

"Proving the Vector Projection Formula," <https://www.youtube.com/watch?v=MdfArHIKWZc>

"Orthogonal Projections," <http://mathonline.wikidot.com/orthogonal-projections>