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

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \alpha_i\left(\nabla f(\boldsymbol{x}_i)\right) \Leftarrow \text{گ‌ن}$$

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i(-\nabla f(\boldsymbol{x_i}) + \beta_i \boldsymbol{p}_i) \Leftarrow \text{گ‌م}$$

روشی مبتنی بر مشتق مرتبه دوم

- ماتریس هسی

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \alpha_i H^{-۱} \nabla f(\boldsymbol{x_i}) \Leftarrow \text{نیوتن}$$

روش‌های تاکنون

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

تاکنون

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

- $V_i = I$
  ؟
- $V_i = H^{-۱}$
  ؟

روش نیوتن

$$f(\boldsymbol{x}_i+\boldsymbol{p}_i)\approx f(\boldsymbol{x}_i)+\nabla f(\boldsymbol{x}_i)^T\boldsymbol{p_i}+\frac{1}{2}\boldsymbol{p_i}^T\nabla^2 f(\boldsymbol{x}_i)\boldsymbol{p_i}$$

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

روش نیوتن: $\boldsymbol{p}$ کاهنده تقریب درجهٔ دو $q(\boldsymbol{p})$

$$\nabla q(\boldsymbol{p_i})=\boldsymbol{g}+H\boldsymbol{p_i}=\boldsymbol{0}$$

معادلهٔ نیوتن

$$H\boldsymbol{p}=-\mathbf{g}$$

$$\boldsymbol{p_i}=-\boldsymbol{H_i^{-1}}\boldsymbol{g_i}$$

$\boldsymbol{p_k}$ جهت نیوتن

روش شبه‌نیوتن

و دیویدن

- آزمایشگاه ملی آرگونه
- در دهه ۳۰ شمسی
- اولین شمارهٔ سیام درباره بهینه‌سازی در سال ۱۳۷۰ شمسی

بین روش تندترین نزول و روش نیوتن

غیرعملی بودن محاسبهٔ ماتریس هسی

تقریب ماتریس وارون هسی

روش شبه‌نیوتن - الگوریتم

مرحله i-اُم

$$\boldsymbol{p_i} = -B_i^{-\mathbf{1}}\boldsymbol{g_i}$$

با جستجو خط (ناکامل) $\boldsymbol{x}_{i+۱} = \boldsymbol{x}_i + \alpha_i\boldsymbol{p_i}$

$$\boldsymbol{g_{i+۱}} = \nabla f(\boldsymbol{x}_{i+۱})$$

$$B_{i+1} \propto (B_i,\boldsymbol{x}_{i+۱}-\boldsymbol{x}_i,\boldsymbol{g_{i+۱}}-\boldsymbol{g_i})$$

روش شبه‌نیوتن

ماتریس هسی

- هزینهٔ بالا یا غیرممکنی محاسبهٔ 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 اندازه‌گیری شده در طول قدم فعلی

روش شبه‌نیوتن

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

B_i تخمینی از $\nabla^2 f(\boldsymbol{x}_i)$

- ماتریس مثبت معین متقارن
- بروز شدن مقدار ماتریس در هر تکرار

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

$\boldsymbol{p} = \boldsymbol{0}$

- $m_i(\boldsymbol{0}) = f_i$
- $\nabla m_i(\boldsymbol{0}) = \nabla f_i$

$m_i(\boldsymbol{p})$ تابع درجهٔ دو کوژ

- کمینه‌ساز تابع
  - $\boldsymbol{p}_i = -B_i^{-۱} \nabla f_i$
- $\boldsymbol{x}_{i+۱} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i$ با ضریبی برآورده‌کنندهٔ شرط‌های وولف
  - مشابه روش نیوتن
    با تخمین به جای مقدار معکوس هسی

روش شبه‌نیوتن

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

ایجاد مدل جدید حول نقطهٔ $\boldsymbol{x}_{i+۱}$ استفاده از

$$m_{i+۱}(\boldsymbol{p}) = f_{i+۱} + \boldsymbol{p}^T \nabla f_{i+۱} + \boldsymbol{p}^T B_{i+۱} \boldsymbol{p}$$

به دنبال $B_{i+۱}$

- بر اساس دانش بدست آمده در مراحل آخر
- برابری گرادیان $m_{i+۱}$ در دو نقطه $\boldsymbol{x}_i$ و $\boldsymbol{x}_{i+۱}$
  - $\nabla m_{i+۱}(\boldsymbol{0}) = \nabla f_{i+۱}$
  - $\nabla m_{i+۱}(-\alpha_i \boldsymbol{p}_i) = \nabla f_{i+۱} - B_{i+۱} \boldsymbol{s}_i$
  - $\nabla m_{i+۱}(-\alpha_i \boldsymbol{p}_i) = \nabla f_i$
  - $\nabla f_i = \nabla f_{i+۱} - B_{i+۱} \boldsymbol{s}_i$

$$B_{i+۱} \boldsymbol{s}_i = \nabla f(\boldsymbol{x}_{i+۱}) - \nabla f(\boldsymbol{x}_i)$$

- معروف به «معادلهٔ وتر»

روش شبه‌نیوتن - ادامه

چگونه؟

- B_i دارای اطلاع
- نو کردن مناسب با استفاده از اطلاعات قبلی

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i$$
$$\boldsymbol{x}_{i+1} - \boldsymbol{x}_i = \alpha_i \boldsymbol{p}_i$$
$$\boldsymbol{s}_i = \boldsymbol{x}_{i+1} - \boldsymbol{x}_i = \alpha_i \boldsymbol{p}_i$$
$$\boldsymbol{g}_i = \nabla f(\boldsymbol{x}_i)$$
$$\boldsymbol{y}_i = \boldsymbol{g}_{i+1} - \boldsymbol{g}_i$$

محاسبهٔ B_{i+1}

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

یا

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

معروف به «معادلهٔ وتر» (نمایش‌های دیگر)

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

روش شبه‌نیوتن - ادامه

هم‌چنین خواهان

- تقارن B_{i+1}
- اختلاف کم B_{i+1} با B_i
- B_i مثبت معین، آن‌گاه B_{i+1} مثبت معین

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i$$
$$\boldsymbol{x}_{i+1} - \boldsymbol{x}_i = \alpha_i \boldsymbol{p}_i$$
$$\boldsymbol{s}_i = \boldsymbol{x}_{i+1} - \boldsymbol{x}_i = \alpha_i \boldsymbol{p}_i$$
$$\boldsymbol{g}_i = \nabla f(\boldsymbol{x}_i)$$
$$\boldsymbol{y}_i = \boldsymbol{g}_{i+1} - \boldsymbol{g}_i$$
$$B_{i+1}\boldsymbol{s}_i = \boldsymbol{y}_i$$

روش شبه‌نیوتن – نوکردن متقارن رتبهٔ-یک

$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i$
$\boldsymbol{x}_{i+1} - \boldsymbol{x}_i = \alpha_i \boldsymbol{p}_i$
$\boldsymbol{s}_i = \boldsymbol{x}_{i+1} - \boldsymbol{x}_i = \alpha_i \boldsymbol{p}_i$
$\boldsymbol{g}_i = \nabla f(\boldsymbol{x}_i)$
$\boldsymbol{y}_i = \boldsymbol{g}_{i+1} - \boldsymbol{g}_i$
$B_{i+1}\boldsymbol{s}_i = \boldsymbol{y}_i$

نو کردن با ماتریس رتبهٔ یک

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

- $a = +1$ یا $a = -1$
- انتخاب a و $\boldsymbol{u}$ به نحوی که برآورده‌ساز معادلهٔ وتر باشد

$$B_{i+1}\boldsymbol{s}_i = \mathbf{y}_i \Longrightarrow B_i\boldsymbol{s}_i + a\boldsymbol{u}\boldsymbol{u}^T\boldsymbol{s}_i = \mathbf{y}_i$$
$$\Longrightarrow (a\boldsymbol{u}^T\boldsymbol{s}_i)\boldsymbol{u} = \mathbf{y}_i - B_i\boldsymbol{s}_i$$

- $\boldsymbol{u}^T\boldsymbol{s}_i$ صرفا ضریب، پس جابجاپذیر

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

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

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

روش شبه‌نیوتن – نوکردن متقارن رتبهٔ-یک

$$\mathbf{y}_i - B_i\boldsymbol{s}_i = a\zeta^2(\mathbf{y}_i - B_i\boldsymbol{s}_i)(\mathbf{y}_i - B_i\boldsymbol{s}_i)^T\boldsymbol{s}_i$$

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

- a
  - برابر علامت $[\boldsymbol{s}_i{}^T(\mathbf{y}_i - B_i\boldsymbol{s}_i)]$
- ζ
  - $\pm[\boldsymbol{s}_i{}^T(\mathbf{y}_i - B_i\boldsymbol{s}_i)]^{-\frac{1}{2}}$

نو کردن با ماتریس رتبهٔ یک

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

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

نتیجه: نوکردن متقارن رتبه-یک برآورده‌کنندهٔ معادلهٔ وتر

$$B_{i+1} = B_i + a\zeta^2(\mathbf{y}_i - B_i\boldsymbol{s}_i)(\mathbf{y}_i - B_i\boldsymbol{s}_i)^T$$

روش شبه‌نیوتن – نوکردن متقارن رتبهٔ-یک

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

- a
  - برابر علامت $[\boldsymbol{s}_i{}^T(\mathbf{y}_i - B_i\boldsymbol{s}_i)]$
- ζ
  - $\pm[\boldsymbol{s}_i{}^T(\mathbf{y}_i - B_i\boldsymbol{s}_i)]^{-\frac{1}{2}}$

نو کردن با ماتریس رتبهٔ یک

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

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

نتیجه: نوکردن متقارن رتبه-یک برآورده‌کنندهٔ معادلهٔ وتر

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

نوکردن متقارن رتبهٔ-یک (مر-۱)-ادامه

در واقع به دنبال معکوس B_{i+1}

نوکردن متقارن رتبهٔ-یک (مر-۱)-ادامه

در واقع به دنبال معکوس B_{i+1}

قضیهٔ معادله شرمن-موریسن-وودباری

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

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

اثبات- تمرین

نوکردن متقارن رتبۀ-یک (مر-۱)-ادامه

در واقع به دنبال معکوس B_{i+1}

تعمیم قضیۀ معادله شرمن-موریسن-وودباری

اگر $A \in \mathbb{R}^{n\times n}$ و $\mathrm{U},V \in \mathbb{R}^{n\times d}$ آن‌گاه $A + UV^T$ معکوس‌پذیر خواهد بود اگر و فقط اگر $I + V^T A^{-1} U$ معکوس‌پذیر باشد. در این‌صورت:

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

نوکردن متقارن رتبۀ-یک (مر-۱)-ادامه

در واقع به دنبال معکوس B_{i+1}

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

قضیۀ معادله شرمن-موریسن-وودباری

اگر $A \in \mathbb{R}^{n\times n}$ و $\boldsymbol{u},\boldsymbol{v} \in \mathbb{R}^n$ و $A + \boldsymbol{u}\boldsymbol{v}^T$ ناتکین‌باشد، آن‌گاه معکوس آن برابر است با

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

آن‌گاه مر-۱ H معکوس B:

$$H_{i+1} = H_i + \frac{(\mathbf{s}_i - H_i\boldsymbol{y}_i)(\mathbf{s}_i - H_i\boldsymbol{y}_i)^T}{(\mathbf{s}_i - H_i\boldsymbol{y}_i)^T\boldsymbol{y}_i}$$

نوکردن متقارن رتبۀ-یک (مر-۱)- ادامه

$$B_{i+۱} = B_i + \frac{(\mathbf{y}_i - B_i\boldsymbol{s}_i)(\mathbf{y}_i - B_i\boldsymbol{s}_i)^T}{(\mathbf{y}_i - B_i\boldsymbol{s}_i)^T\boldsymbol{s}_i}$$

آن‌گاه نوکردن مر-۱ H معکوس B:

$$H_{i+۱} = H_i + \frac{(\mathbf{s}_i - H_i\boldsymbol{y}_i)(\mathbf{s}_i - H_i\boldsymbol{y}_i)^T}{(s - H_i\boldsymbol{y}_i)^T\boldsymbol{y}_i}$$

نوکردن متقارن رتبهٔ-یک (مر-۱)- ادامه

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

نوکردن متقارن رتبهٔ-یک (مر-۱)-ادامه

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

عدم رعایت مثبت معین بودن

$$H_{i+۱} = H_i + \frac{(\mathbf{s}_i - H_i \boldsymbol{y}_i)(\mathbf{s}_i - H_i \boldsymbol{y}_i)^T}{(\mathbf{s}_i - H_i \boldsymbol{y}_i)^T \boldsymbol{y}_i}$$

روش‌های نوکردن تکراری شبه‌نیوتن دیگر

مر-۱

$$H_{i+۱} = H_i + \frac{(\mathbf{s}_i - H_i\boldsymbol{y}_i)(\mathbf{s}_i - H_i\boldsymbol{y}_i)^T}{(\mathbf{s}_i - H_i\boldsymbol{y}_i)^T\boldsymbol{y}_i}$$

روش‌های نوکردن تکراری شبه‌نیوتن دیگر

دیویدن-فلچر-پاول DFP

$$H_{i+۱} = H_i - \frac{H_i \boldsymbol{y}_i \boldsymbol{y}_i^T H_i}{\boldsymbol{y}_i^T H_i \boldsymbol{y}_i} + \frac{\boldsymbol{s}_i \boldsymbol{s}_i^T}{\boldsymbol{y}_i^T \boldsymbol{s}_i}$$

روش‌های نوکردن تکراری شبه‌نیوتن دیگر

برویدن–فلچر–گولدفرب–شانو BFGS

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

$$H_{i+۱} = (I - \gamma_i \boldsymbol{s}_i \boldsymbol{y}_i{}^T) H_i (I - \gamma_i \mathbf{y}_i \boldsymbol{s}_i{}^T) + \gamma_i \mathbf{s}_i \boldsymbol{s}_i{}^T$$

روش‌های نوکردن تکراری شبه‌نیوتن دیگر

بروید‌ن-فلچر-گولدفرب-شانو BFGS

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

$$H_{i+١} = (I - \gamma_i \mathbf{s}_i \boldsymbol{y}_i{}^T) H_i (I - \gamma_i \mathbf{y}_i \boldsymbol{s}_i{}^T) + \gamma_i \mathbf{s}_i \boldsymbol{s}_i{}^T$$

$\boldsymbol{s}_i$ یا $\boldsymbol{p}_i$

مقداردهی اولیه

- $B_0 = I$
- $B_0 = H^{-1}$

اگر تابع درجه دو باشد

- ماتریس هسی دقیق بعد از **n** مرحله به دست خواهد آمد
- متناظر با گم زمانی که $B_0 = I$

اگر تابع غیردرجه دو

- شن معمولا سریعتر از گم
- عدم نیاز به جستجو خط کامل

الگوریتم

انتخاب $\boldsymbol{x}_0$ و ϵ

محاسبه ماتریس وارون هسی H_0

$i = 0$

تازمان $\|\nabla f_i\| > \epsilon$

- محاسبهٔ جهت جستجو

$$\boldsymbol{p}_i = -H_i \nabla f_i$$

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

یافتن α_i با جستجوی خط تا یافتن مقدار برآورده‌کننده شرط وولف

$$\boldsymbol{s}_i = \boldsymbol{x}_{i+1} - \boldsymbol{x}_i$$

$$\boldsymbol{y}_i = \nabla \mathrm{f}_{i+1} - \nabla \mathrm{f}_i$$

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

$$H_{i+1} = (I - \gamma_i \mathbf{s}_i \boldsymbol{y}_i{}^T) H_i (I - \gamma_i \mathbf{y}_i \boldsymbol{s}_i{}^T) + \gamma_i \mathbf{s}_i \boldsymbol{s}_i{}^T$$

$$i = i + 1$$

خانواده بریدون

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

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

برویدن-فلچر-گولدفرب-شانو

دیویدن-فلچر-پاول DFP

خانواده بریدون

$$B_{i+۱} = B_i - \frac{B_i \boldsymbol{s}_i \boldsymbol{s}_i^T B_i}{\boldsymbol{s}_i^T B_i \boldsymbol{s}_i} + \frac{\boldsymbol{y}_i \boldsymbol{y}_i^T}{\boldsymbol{y}_i^T \boldsymbol{s}_i} + \phi_i (\boldsymbol{s}_i^T B_i \boldsymbol{s}_i) \boldsymbol{v}_i \boldsymbol{v}_i^T$$

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

برویدن-فلچر-گولدفرب-شانو
- $\phi_i = 0$

دیویدن-فلچر-پاول DFP
- $\phi_i = ۱$

خانواده بریدون

ترکیب

$$B_{i+۱} = (۱ - \phi_i) B_{i+۱}^{BFGS} + \phi_i B_{i+۱}^{DFP}$$

برویدن-فلچر-گولدفرب-شانو

- $\phi_i = 0$

داویدون-فلچر-پاول DFP

- $\phi_i = ۱$

شبه‌نیوتن

هزینهٔ بالا یا غیرممکنی محاسبهٔ 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
  - $O(n^۲)$ ضرب به ازای هر مرحله

معایب

- امکان نبود جهت نزول
- امکان دقیق‌نبودن تخمین هسی

منابع

[نازه‌دل]

[الوئن‌برگر]

[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=MdfArHlKWZc

“Orthogonal Projections,” http://mathonline.wikidot.com/orthogonal-projections