



بهینه سازی
مبانی بهینه سازی نامقید
ذیل

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

روش‌های جستجو خط

- یافتن مسیر جستجو p_i
- تصمیم بر اندازه قدم در راستای مسیر جستجو

$$x_{i+1} = x_i + \alpha_i p_i$$

α_i ▪

- طول قدم
- سرعت یادگیری

p_i ▪

- تندترین نزول
- مزدوج گرادیان
- نیوتن
- شبه‌نیوتن

روش‌های جستجو خط

- یافتن مسیر جستجو \mathbf{p}_i
- تصمیم بر اندازه قدم در راستای مسیر جستجو

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

▪ α_i

- طول قدم
- سرعت یادگیری

▪ \mathbf{p}_i

- تندترین نزول

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

$$\mathbf{p}_i = -I \nabla f(\mathbf{x}_i)$$

- I ماتریس همانی

- مزدوج گرادیان

- نیوتن

- شبه‌نیوتن

روش‌های جستجو خط

- یافتن مسیر جستجو \mathbf{p}_i
- تصمیم بر اندازه قدم در راستای مسیر جستجو

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

- α_i
- طول قدم
- سرعت یادگیری

▪ \mathbf{p}_i

- تندترین نزول
- مزدوج گرادیان

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

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

- فلچر-ریوز

$$\beta_i = \frac{\|\mathbf{g}_{i+1}\|_2^2}{\|\mathbf{g}_i\|_2^2}$$

- پولاک-ریبرییه

$$\beta_i = \frac{\mathbf{g}_{i+1}^T (\mathbf{g}_{i+1} - \mathbf{g}_i)}{\|\mathbf{g}_i\|_2^2}$$

- نیوتن
- شبه نیوتن

روش‌های جستجو خط

- یافتن مسیر جستجو \mathbf{p}_i
- تصمیم بر اندازه قدم در راستای مسیر جستجو

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

▪ α_i

- طول قدم
- سرعت یادگیری

▪ \mathbf{p}_i

- تندترین نزول
- مزدوج گرادپان
- نیوتن

$$\mathbf{p}_i = -H^{-1} \nabla f(\mathbf{x}_i) \quad \text{▪}$$

$$H = \nabla^2 f(\mathbf{x}_i) \quad \text{▪}$$

- نیوتن تغییر یافته

$$\mathbf{p}_i = -(H + \nu I)^{-1} \nabla f(\mathbf{x}_i) \quad \text{▪}$$

- شبه نیوتن

روش‌های جستجو خط

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

- یافتن مسیر جستجو \mathbf{p}_i
- تصمیم بر اندازه قدم در راستای مسیر جستجو

α_i

- طول قدم
- سرعت یادگیری

\mathbf{p}_i

- تندترین نزول
- مزدوج گرادینان
- نیوتن
- شبه‌نیوتن

$$\mathbf{p}_i = -B_i^{-1} \nabla f(\mathbf{x}_i)$$

- B_i تقریب ماتریس هسی
- با استفاده از ماتریس رتبه پایین

▪ مر-۱

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

▪ دلف

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

▪ ب‌ف‌گ‌ش

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

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

روش‌های جستجو خط

- یافتن مسیر جستجو \mathbf{p}_i
- تصمیم بر اندازه قدم در راستای مسیر جستجو

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

- α_i
- طول قدم
- سرعت یادگیری

- \mathbf{p}_i
- تندترین نزول
- مزدوج گرادپان
- نیوتن
- شبه نیوتن
- همه ماتریس‌ها متقارن و ناتکین
- مثبت معین

$$\mathbf{p}_i^T \nabla f(\mathbf{x}_i) = -\nabla f(\mathbf{x}_i)^T B_i^{-1} \nabla f(\mathbf{x}_i) < 0$$

پس \mathbf{p}_i مسیری نزولی

همگرایی روش های جستجو خط

الگوریتم های عددی تکراری (مرحله به مرحله)

بهینه سازی مناسب بسته به انتخاب طول قدم مناسب و جهت مناسب

همگرایی به جهت می پردازد

از زاویه دیگر

▪ همگرایی به معنی رسیدن به نقطه کمینه

همگرایی سراسری:

▪ الگوریتم با شروع از هر فاصله دوری از کمینه، به کمینه برسد

▪ همگرایی سراسری روش های جستجو خطی $x_{i+1} = x_i + \alpha_i p_i$

▪ به نقطه مانا

▪ مگر استفاده از اطلاعات هسی

مرتبه همگرایی

کوچکترین کران بالای اعداد نامنفی p صادق در

$$0 \leq \lim_{i \rightarrow \infty} \frac{|r_{i+1} - r^*|}{|r_i - r^*|^p} < \infty$$

ملاک بررسی در حین میل به بی‌نهایت
▪ ذیل مقادیر و مراحل

کاهش فاصله از حد به اندازه توان p

مرتبہ همگرائی

فرض وجود دنباله و حد زیر

$$\beta = \lim_{i \rightarrow \infty} \frac{|r_{i+1} - r^*|}{|r_i - r^*|^p}$$

آن گاه

$$|r_{i+1} - r^*| = \beta |r_i - r^*|^p$$

$$\beta = \lim_{i \rightarrow \infty} \frac{|r_{i+1} - r^*|}{|r_i - r^*|^p}$$

مرتبہ همگرائی - مثال

$$0 < a < 1$$

$$r_i = a^i \text{ (الف)}$$

▪ همگرا به صفر با مرتبہ یک

$$\frac{r_{i+1}}{r_i} = a \text{ ▪}$$

$$r_i = a^{(2^i - 1)} \text{ (ب)}$$

▪ همگرا به صفر با مرتبہ دو

$$\frac{r_{i+1}}{r_i^2} = a \text{ ▪}$$

همگرایی خطی

$$p = 1$$

$$\lim_{i \rightarrow \infty} \frac{|r_{i+1} - r^*|}{|r_i - r^*|} = \beta < 1$$

همگرا به حد با نسبت (سرعت) همگرایی β

در i گام: سرعت $c\beta^i$

▪ همگرایی هندسی!

مثال - مقایسه دو الگوریتم خطی

▪ مقایسه بر حسب نسبت‌های همگرایی متناظر آنها

▪ نسبت کوچکتر نمایشگر سرعت بیشتر

$\beta = 0$ همگرایی زبرخطی

همگرایی روش‌های جستجو خط

تندترین نزول همیشه در راستای کاهش گرادیان پس همگرا به نقطه مانا

روش نیوتن-محور

- در صورتی که ماتریس هسی یا تقریب هسی
- دارای عدد شرط حددار
- مثبت معین
- رعایت شروط وولف

روش گرادیان مزدوج

$$Qx = b$$

- همگرایی در صورت تقارن و مثبت معینی در بی‌نهایت
- بسته به عدد وضعیت

سرعت همگرایی

همگرایی آسان به دنبال مسیرهای غیرمتعامل با گرادیان

- ولی کافی نیست

- گاهی اوقات موجب اشتباه

بسیار کند

- تندترین نزول

روش نیوتن رسیدن به جواب وقتی در همسایگی باشد

- ولی در دور دست لزوماً به جواب نخواهد رسید

سرعت همگرایی گرادیان نزولی

تحلیل تابع درجه دو

جستجو خط کامل

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T Q \mathbf{x} - \mathbf{b}^T \mathbf{x}$$

Q متقارن و مثبت معین

گرادیان

$$\nabla f(\mathbf{x}) = Q \mathbf{x} - \mathbf{b}$$

کمینه‌ساز پاسخ منحصر به فرد $Q \mathbf{x} = \mathbf{b}$

سرعت همگرایی گرادیان نزولی

محاسبه طول قدم کمینه‌ساز تابع $f(\mathbf{x}_i - \alpha \nabla f_i)$

$$f(\mathbf{x}_i - \alpha \nabla f_i) = \frac{1}{2} (\mathbf{x}_i - \alpha \nabla f_i)^T Q (\mathbf{x}_i - \alpha \nabla f_i) - \mathbf{b}^T (\mathbf{x}_i - \alpha \nabla f_i)$$

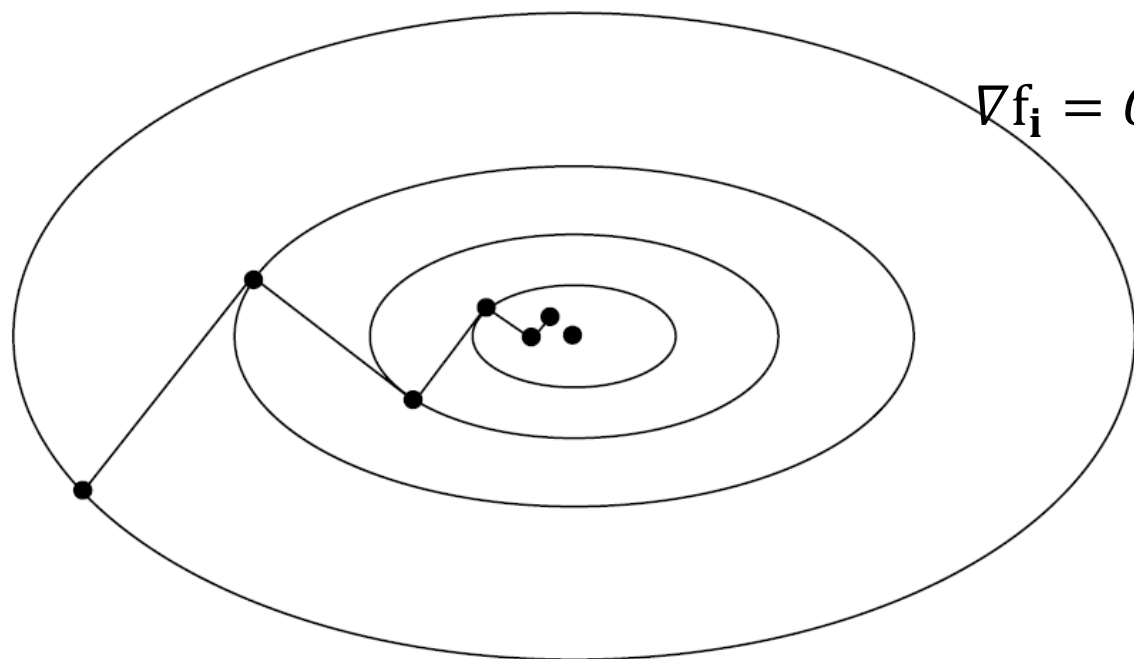
با مشتق‌گیری نسبت به طول قدم و برابر صفر قرار دادن

$$\alpha_i = \frac{\nabla f_i^T \nabla f_i}{\nabla f_i^T Q \nabla f_i}$$

سرعت همگرایی گرادینان نزولی

با استفاده از مقدار دقیق کمینه‌ساز طول قدم

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \left(\frac{\nabla f_i^T \nabla f_i}{\nabla f_i^T Q \nabla f_i} \right) \nabla f_i$$



$$\nabla f_i = Q \mathbf{x}_i - \mathbf{b}$$

امکان نوشتن صورت بسته \mathbf{x}_{i+1} بر اساس \mathbf{x}_i

بیضی‌وارهای n-بعدی

▪ محورهای آن برابر با بردار ویژه‌های ماتریس Q

سرعت همگرایی گرادیان نزولی

$$\|x\|_Q^2 = x^T Q x$$

استفاده از

$$Q x^* = b$$

$$\frac{1}{2} \|x - x^*\|_Q^2 = f(x) - f(x^*)$$

امکان استفاده از سمت چپ معادله برای تقریب سمت راست و استفاده از آن در همگرایی

سرعت همگرایی گرادیان نزولی

$$\nabla f_i = Q(\mathbf{x}_i - \mathbf{x}^*)$$

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \left(\frac{\nabla f_i^T \nabla f_i}{\nabla f_i^T Q \nabla f_i} \right) \nabla f_i$$

$$\|\mathbf{x}_{i+1} - \mathbf{x}^*\|_Q^2 = \left\{ 1 - \frac{(\nabla f_i^T \nabla f_i)^2}{(\nabla f_i^T Q \nabla f_i)(\nabla f_i^T Q^{-1} \nabla f_i)} \right\} \|\mathbf{x}_i - \mathbf{x}^*\|_Q^2$$

مقدار دقیق کاهش f در هر مرحله تکرار الگوریتم

▪ سختی تفسیر

سرعت همگرایی گرادیان نزولی

استفاده از عدد وضعیت (شرط)

قضیه

روش تندترین نزول با جستجو خط دقیق روی تابع درجه دو محدب

$$\|x_{i+1} - x^*\|_Q^2 = \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^2 \|x_i - x^*\|_Q^2$$

بردارهای ویژه Q

$$0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$$

$$\kappa(Q) = \|Q\| \|Q^{-1}\| \cdot$$

• $\kappa(Q)$ کوچک: «خوش وضعیتی»

• $\kappa(Q)$ بزرگ: بدوضعیتی!

• گاهی اوقات امکان تعریف $\kappa(Q)$ با مقادیر ویژه

• چه مواقعی؟

سرعت همگرایی گرادیان نزولی

$$\|x_{i+1} - x^*\|_Q^2 = \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^2 \|x_i - x^*\|_Q^2$$

عدد شرط (همیشه لزوما این گونه نیست)

$$\kappa(Q) = \|Q\| \|Q^{-1}\| = \frac{\lambda_n}{\lambda_1}$$

$$\|x_i - x^*\|_Q = \left(\frac{\kappa(Q) - 1}{\kappa(Q) + 1} \right)^i \|x_0 - x^*\|_Q$$

- افزایش عدد شرط
- کشیده شدن سطح ترازها
- افزایش زیگزاگها
- نزول سرعت همگرایی
- عدد شرط بزرگ برابر با ماتریس بدتعریف

سرعت همگرایی گرادیان نزولی

$$r \in \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}, 1 \right)$$

$$f(x_{i+1}) - f(x^*) \leq r[f(x_i) - f(x^*)]$$

مثال

▪ $f(x^*) = 0, f(x_1) = 1, \kappa(Q) = 100$

▪ آنگاه با تقریب بالا

▪ پس از هزار قدم

▪ $f(x_{1000}) = 0.08$

همگرایی روش نیوتن

همگرایی محلی

مجانبی درجه دو

با فرض

$$H(\mathbf{x}^*) > 0 \quad \blacksquare$$

▪ پیوسته و هموار

انواع روش‌های محاسبه خطا

$$\delta_i = \|f(\mathbf{x}_i) - f(\mathbf{x}^*)\| \text{ یا } \delta_i = \|\mathbf{x}_i - \mathbf{x}^*\|$$

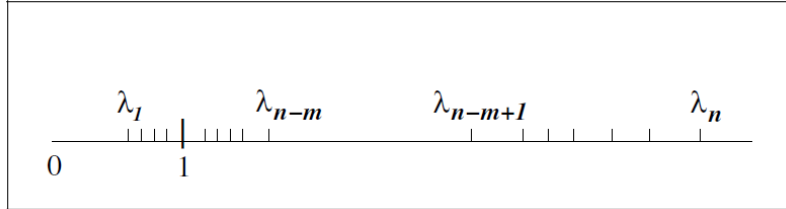
$$\delta_{i+1} \leq c\delta_i^2$$

$$\text{مثال } \delta_k = 10^{-2} \text{ و } \delta_{i+1} = 10^{-4} \text{ و } \delta_{i+2} = 10^{-8} \quad \blacksquare$$

سرعت همگرایی گرادیان مزدوج

$$\|x_{i+1} - x^*\|_Q^2 \leq \left(\frac{\lambda_{n-i} - \lambda_1}{\lambda_{n-i} + \lambda_1} \right)^2 \|x_i - x^*\|_Q^2$$

سرعت همگرایی گرادیان مزدوج

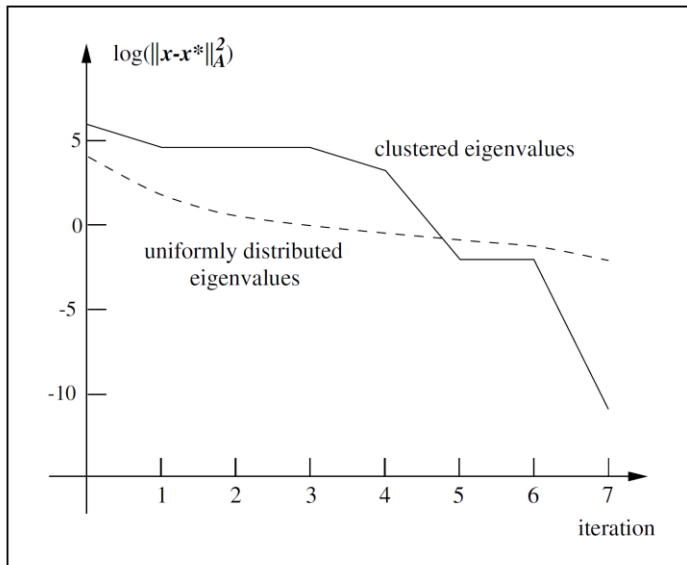


- در صورتی که Q مثبت معین و متقارن با اندازه $n \times n$
- با مقادیر ویژه $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$
 - دارای m مقدار ویژه بزرگ و $n - m$ مقدار ویژه کوچک حول یک
 - پس از $m + 1$ قدم، داریم $\epsilon = \lambda_{n-m} - \lambda_1$

$$\|x_{m+1} - x^*\|_Q \approx \epsilon \|x_0 - x^*\|_Q$$

به چه معنا

- رسیدن به تقریب خوب با $m + 1$ قدم



• مثال - دو خوشه

• خوشه مقادیر ویژه بزرگ دارای پنج عضو

سرعت همگرایی گرادیان مزدوج

تقریب بیشتر

$$\kappa(Q) = \|Q\| \|Q^{-1}\| = \frac{\lambda_n}{\lambda_1}$$

$$\|x_i - x^*\|_Q \leq \left(\frac{\sqrt{\kappa(Q)} - 1}{\sqrt{\kappa(Q)} + 1} \right)^i \|x_0 - x^*\|_Q$$

مناسب در مواقع داشتن تقریب بالا و پایین مقدار ویژه‌ها

▪ مقایسه با گرادیان نزولی

$$\|x_i - x^*\|_Q = \left(\frac{\kappa(Q) - 1}{\kappa(Q) + 1} \right)^i \|x_0 - x^*\|_Q \quad \square$$

سرعت همگرایی گرادیان مزدوج

$$0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$$

$$\|\mathbf{x}_{i+1} - \mathbf{x}^*\|_Q^2 \leq \left(\frac{\lambda_{n-i} - \lambda_1}{\lambda_{n-i} + \lambda_1} \right)^2 \|\mathbf{x}_i - \mathbf{x}^*\|_Q^2$$

$$\|\mathbf{x}_{i+1} - \mathbf{x}^*\|_Q \approx \epsilon \|\mathbf{x}_0 - \mathbf{x}^*\|_Q$$

$$\|\mathbf{x}_i - \mathbf{x}^*\|_Q \leq 2 \left(\frac{\sqrt{\kappa(Q)} - 1}{\sqrt{\kappa(Q)} + 1} \right)^i \|\mathbf{x}_0 - \mathbf{x}^*\|_Q$$

- $\kappa(Q) = \|Q\| \|Q^{-1}\|$
- $\kappa(Q)$ کوچک: «خوش وضعیتی»
- $\kappa(Q)$ بزرگ: بدوضعیتی!»
- گاهی اوقات امکان تعریف $\kappa(Q)$ با مقادیر ویژه
- چه مواقعی؟

سرعت همگرایی گرادیان مزدوج

$$\|x_{i+1} - x^*\| \leq \frac{\sqrt{\lambda_n} - \sqrt{\lambda_1}}{\sqrt{\lambda_n} + \sqrt{\lambda_1}} \|x_i - x^*\|$$

در صورت غلبه λ_1 بر λ_n

$$\frac{\sqrt{\lambda_n} - \sqrt{\lambda_1}}{\sqrt{\lambda_n} + \sqrt{\lambda_1}} = 1 - 2\sqrt{\frac{\lambda_1}{\lambda_n}} = 1 - \frac{2}{\sqrt{\kappa(Q)}}$$

تابع درجه دو
▪ بیضی‌های متحدالمركز

هر چه $\kappa(Q)$ بزرگتر
▪ بیضی کشیده‌تر
▪ کاهش همگرایی

منابع

[نازهدل]

[لوئینبرگر]