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 $\boldsymbol{p}_i$
- تصمیم بر اندازه قدم در راستای مسیر جستجو

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

- α_i
  - طول قدم
  - سرعت یادگیری
- $\boldsymbol{p}_i$
  - تندترین نزول
  - مزدوج گرادیان
  - نیوتن
  - شبه‌نیوتن

روش‌های جستجو خط

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

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

- α_i
  - طول قدم
  - سرعت یادگیری
- $\boldsymbol{p}_i$
  - تندترین نزول
    - $\boldsymbol{p}_i = -\nabla f(\boldsymbol{x}_i)$
    - $\boldsymbol{p}_i = -\mathrm{I}\nabla f(\boldsymbol{x}_i)$
    - I ماتریس همانی
  - مزدوج گرادیان
  - نیوتن
  - شبه‌نیوتن

روش‌های جستجو خط

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

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

- α_i
  - طول قدم
  - سرعت یادگیری
- $\boldsymbol{p}_i$
  - تندترین نزول
  - مزدوج گرادیان
  - $\boldsymbol{g}_i = \nabla f(\boldsymbol{x}_i)$
    - $\boldsymbol{p}_i = -\boldsymbol{g}_i + \boldsymbol{\beta}_i \boldsymbol{p}_{i-1}$
      - فلچر-ریوز
      - $\beta_i = \frac{\|\boldsymbol{g}_{i+1}\|_2^2}{\|\boldsymbol{g}_i\|_2^2}$
      - پولاک-ریبریه
      - $\beta_i = \frac{\boldsymbol{g}_{i+1}{}^T(\boldsymbol{g}_{i+1} - \boldsymbol{g}_i)}{\|\boldsymbol{g}_i\|_2^2}$
  - نیوتن
  - شبه‌نیوتن

روش‌های جستجو خط

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

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

- α_i
  - طول قدم
  - سرعت یادگیری
- $\boldsymbol{p}_i$
  - تندترین نزول
  - مزدوج گرادیان
  - نیوتن
    - $\boldsymbol{p}_i = -H^{-1} \nabla f(\boldsymbol{x}_i)$
      - $H = \nabla^2 f(\boldsymbol{x}_i)$
    - نیوتن تغییریافته
      - $\boldsymbol{p}_i = -(H + \nu I)^{-1} \nabla f(\boldsymbol{x}_i)$
  - شبه‌نیوتن

روش‌های جستجو خط

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

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

- α_i
  - طول قدم
  - سرعت یادگیری
- $\boldsymbol{p}_i$
  - تندترین نزول
  - مزدوج گرادیان
  - نیوتن
  - شبه‌نیوتن
    - $\boldsymbol{p}_i = -B_i^{\text{م}} \nabla f(\boldsymbol{x}_i)$
    - B_i تقریب ماتریس هسی
    - با استفاده از ماتریس رتبه پایین
      - مر-۱
      - $B_{i+1}^{\text{م}} = B_i^{\text{م}} + \frac{(\mathbf{s}_i - B_i^{\text{م}} \boldsymbol{y}_i)(\mathbf{s}_i - B_i^{\text{م}} \boldsymbol{y}_i)^T}{(\mathbf{s}_i - B_i^{\text{م}} \boldsymbol{y}_i)^T \boldsymbol{y}_i}$
      - دف‌پ
      - $B_{i+1}^{\text{م}} = B_i^{\text{م}} - \frac{B_i^{\text{م}} \boldsymbol{y}_i \boldsymbol{y}_i^T B_i^{\text{م}}}{\boldsymbol{y}_i^T B_i^{\text{م}} \boldsymbol{y}_i} + \frac{\mathbf{s}_i \mathbf{s}_i^T}{\boldsymbol{y}_i^T \mathbf{s}_i}$
      - ب‌ف‌گ‌ش
      - $B_{i+1}^{\text{م}} = (I - \gamma_i \mathbf{s}_i \boldsymbol{y}_i^T) B_i^{\text{م}} (I - \gamma_i \mathbf{y}_i \mathbf{s}_i^T) + \gamma_i \mathbf{s}_i \mathbf{s}_i^T$

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

روش‌های جستجو خط

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

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

- α_i
  - طول قدم
  - سرعت یادگیری
- $\boldsymbol{p}_i$
  - تندترین نزول
  - مزدوج گرادیان
  - نیوتن
  - شبه‌نیوتن
- همه ماتریس‌ها متقارن و ناتکین
- مثبت معین

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

روش‌های جستجو خط

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

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

- α_i
  - طول قدم
  - سرعت یادگیری
- $\boldsymbol{p}_i$
  - تندترین نزول
  - مزدوج گرادیان
  - نیوتن
  - شبه‌نیوتن
- همه ماتریس‌ها متقارن و ناتکین
- مثبت معین

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

پس $\boldsymbol{p}_i$ مسیری نزولی

همگرائی روش‌های جستجو خط

الگوریتم‌های عددی تکراری (مرحله‌به‌مرحله)

بهینه‌سازی مناسب بسته به انتخاب طول قدم مناسب و جهت مناسب

همگرائی به جهت می‌پردازد

از زاویه دیگر

- همگرائی به معنی رسیدن به نقطه کمینه

همگرائی سراسری:

- الگوریتم با شروع از هر فاصلهٔ دوری از کمینه، به کمینه برسد

- همگرایی سراسری روش‌های جستجو خطی $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i$
  - به نقطه مانا
  - مگر استفاده از اطلاعات هسی

مرتبه همگرائی

کوچکترین کران بالای اعداد نامنفی p صادق در

$$0 \le \underset{i\to\infty}{\overline{\text{حد}}} \frac{|r_{i+۱} - r^*|}{|r_i - r^*|^p} < \infty$$

ملاک بررسی در حین میل به بی‌نهایت

- ذیل مقادیر و مراحل

کاهش فاصله از حد به اندازه توان p

مرتبه همگرائی

فرض وجود دنباله و حد زیر

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

آن‌گاه

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

مرتبه همگرائی- مثال

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

$0 < a < 1$

الف) $r_i = a^i$

- همگرا به صفر با مرتبهٔ یک
- $\frac{r_{i+1}}{r_i} = a$

ب) $r_i = a^{(2^i)}$

- همگرا به صفر با مرتبه دو
- $\frac{r_{i+1}}{{r_i}^2} = a$

همگرائی خطی

$p = ۱$

$$\underset{i\to\infty}{حد} \frac{|r_{i+۱} - r^*|}{|r_i - r^*|} = \beta < 1$$

همگرا به حد با نسبت (سرعت) همگرایی β

در i گام: سرعت $c\beta^i$

- هم‌گرائی هندسی!

مثال- مقایسه دو الگوریتم خطی

- مقایسه بر حسب نسبت‌های همگرائی متناظر آن‌ها
- نسبت کوچکتر نمایشگر سرعت بیشتر

$\beta = 0$ همگرائی زبرخطی

همگرائی 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 همگرائی

همگرائی آسان به دنبال مسیرهای غیرمتعامد 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 لزوما به جواب نخواهد رسید

سرعت همگرایی گرادیان نزولی

تحلیل تابع درجه دو

جستجو خط کامل

$$f(\boldsymbol{x}) = \frac{۱}{۲}\boldsymbol{x}^T Q \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x}$$

Q متقارن و مثبت معین

گرادیان

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

کمینه‌ساز پاسخ منحصربه‌فرد $Q\boldsymbol{x} = \boldsymbol{b}$

سرعت همگرایی گرادیان نزولی

محاسبهٔ طول قدم کمینه‌ساز تابع $f(\boldsymbol{x}_i - \alpha\nabla \mathrm{f}_{\boldsymbol{i}})$

$$f(\boldsymbol{x_i} - \alpha\nabla \mathrm{f}_{\boldsymbol{i}}) = \frac{۱}{۲}(\boldsymbol{x_i} - \alpha\nabla \mathrm{f}_{\mathbf{i}})^T Q(\boldsymbol{x_i} - \alpha\nabla \mathrm{f}_{\boldsymbol{i}}) - \boldsymbol{b}^T(\boldsymbol{x_i} - \alpha\nabla \mathrm{f}_{\mathbf{i}})$$

با مشتق‌گیری نسبت به طول قدم و برابر صفر قرار دادن

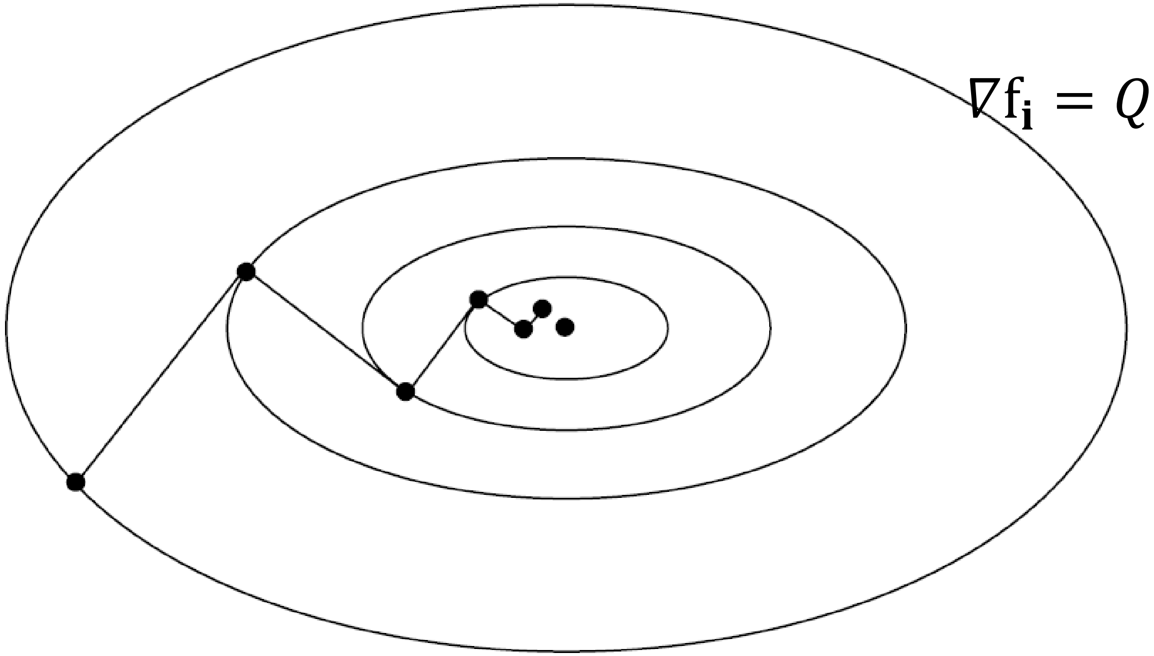
$$\alpha_i = \frac{\nabla \mathrm{f}_{\mathrm{i}}^T \nabla \mathrm{f}_i}{\nabla \mathrm{f}_{\mathrm{i}}^T \mathrm{Q} \nabla \mathrm{f}_i}$$

سرعت همگرایی گرادیان نزولی

با استفاده از مقدار دقیق کمینه‌ساز طول قدم

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

$$\nabla \mathrm{f}_\mathbf{i} = Q\boldsymbol{x}_{\boldsymbol{i}} - \boldsymbol{b}$$



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

بیضی‌وارهای **n**-بعدی

- محورهای آن برابر با بردار ویژه‌های ماتریس Q

سرعت همگرایی گرادیان نزولی

$$\|\boldsymbol{x}\|_Q^2 = \boldsymbol{x}^T Q \boldsymbol{x}$$

استفاده از

$$Q\boldsymbol{x}^* = \boldsymbol{b}$$

$$\frac{۱}{۲}\|\boldsymbol{x} - \boldsymbol{x}^*\|_Q^2 = f(\boldsymbol{x}) - f(\boldsymbol{x}^*)$$

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

سرعت همگرایی گرادیان نزولی

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

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

$$\|\boldsymbol{x}_{i+1} - \boldsymbol{x}^*\|_Q^2 = \left\{ ۱ - \frac{\left(\nabla \mathrm{f_i}^T \nabla \mathrm{f_i}\right)^۲}{\left(\nabla \mathrm{f_i}^T \mathrm{Q} \nabla \mathrm{f_i}\right)\left(\nabla \mathrm{f_i}^T \mathrm{Q}^{-۱} \nabla \mathrm{f_i}\right)} \right\} \|\boldsymbol{x}_{\boldsymbol{i}} - \boldsymbol{x}^*\|_Q^2$$

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

$$\|\boldsymbol{x}_{i+1} - \boldsymbol{x}^*\|_Q^2 = \left(\frac{\lambda_n - \lambda_۱}{\lambda_n + \lambda_۱}\right)^۲ \|\boldsymbol{x_i} - \boldsymbol{x}^*\|_Q^2$$

بردارهای ویژه Q

$$0 \leq \lambda_۱ \leq \lambda_۲ \leq \cdots \leq \lambda_n$$

سرعت همگرایی گرادیان نزولی

- $\kappa(Q) = \|Q\| \left\|Q^{-۱}\right\|$
- $\kappa(Q)$ کوچک: «خوش وضعیتی»
- $\kappa(Q)$ بزرگ: بدوضعیتی!
- گاهی اوقات امکان تعریف $\kappa(Q)$ با مقادیر ویژه
- چه مواقعی؟

$$\|\boldsymbol{x}_{i+1} - \boldsymbol{x}^*\|_Q^2 = \left(\frac{\lambda_n - \lambda_۱}{\lambda_n + \lambda_۱}\right)^۲ \|\boldsymbol{x}_{\boldsymbol{i}} - \boldsymbol{x}^*\|_Q^2$$

عدد شرط (همیشه لزوما این‌گونه نیست)

$$\kappa(Q) = \|Q\|\|Q^{-۱}\| = \frac{\lambda_n}{\lambda_۱}$$

$$\|\boldsymbol{x}_{\boldsymbol{i}} - \boldsymbol{x}^*\|_Q = \left(\frac{\kappa(Q) - ۱}{\kappa(Q) + ۱}\right)^k \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_Q$$

- افزایش عدد شرط
  - کشیده‌شدن سطح‌ترازها
  - افزایش زیگزاگ‌ها
  - نزول سرعت همگرایی
  - عدد شرط بزرگ برابر با ماتریس بدتعریف

سرعت همگرایی گرادیان نزولی

$$r \in \left(\frac{\lambda_n - \lambda_۱}{\lambda_n + \lambda_۱}, ۱\right)$$

$$f(\boldsymbol{x}_{i+۱}) - f(\boldsymbol{x}^*) \leq r[f(\boldsymbol{x}_i) - f(\boldsymbol{x}^*)]$$

مثال

- $\kappa(Q) = ۸۰۰$، $f(x_۱) = ۱$، $f(x^*) = 0$
- آنگاه با تقریب بالا
  - پس از هزار قدم
  - $f(x_{۱...}) = ۰.۰۸$

همگرایی روش نیوتن

همگرایی محلی

مجانبی درجه دو

با فرض

- $H(\boldsymbol{x}^*) \succ 0$
- پیوسته و هموار

انواع روش‌های محاسبه خطا

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

$$\delta_{i+۱} \leq c\delta_i^۲$$

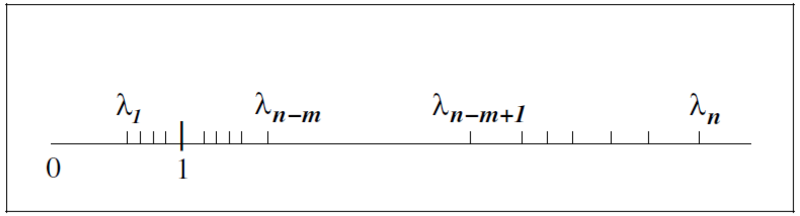
مثال $\delta_k = ۱۰^{-۲}$

- $\delta_{i+۱} = ۱۰^{-۴}$ و $\delta_{i+۲} = ۱۰^{-۸}$

سرعت همگرائی گرادیان مزدوج

$$\|\boldsymbol{x}_{i+۱} - \boldsymbol{x}^*\|_Q^۲ \leq \left(\frac{\lambda_{n-i} - \lambda_۱}{\lambda_{n-i} + \lambda_۱}\right)^۲ \|\boldsymbol{x}_i - \boldsymbol{x}^*\|_Q^۲$$

سرعت همگرائی گرادیان مزدوج



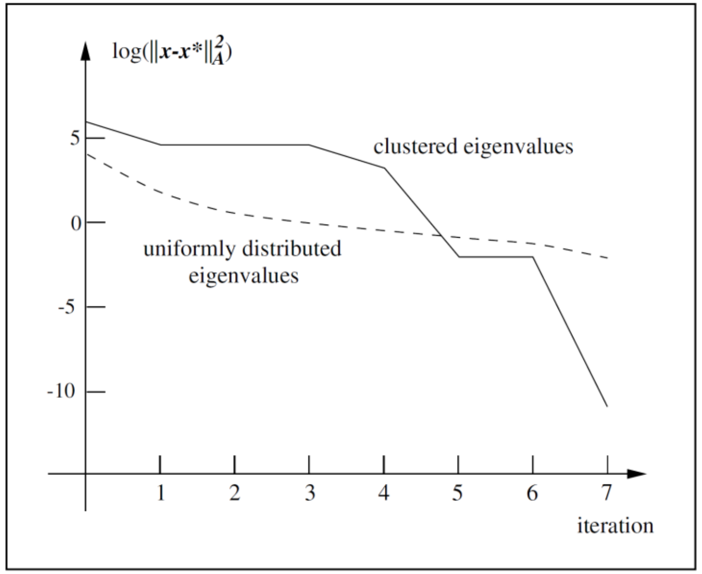
در صورتی که Q مثبت معین و متقارن با اندازهٔ $n \times n$

- با مقادیر ویژهٔ $0 \le \lambda_۱ \le \lambda_۲ \le \cdots \le \lambda_n$
- دارای m مقدار ویژهٔ بزرگ و $n-m$ مقدار ویژهٔ کوچک حول یک
- $\epsilon = \lambda_{n-m} - \lambda_۱$، پس از $m+۱$ قدم،داریم

$$\|\boldsymbol{x}_{m+۱} - \boldsymbol{x}^*\|_Q \approx \epsilon \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_Q$$

به چه معنا

- رسیدن به تقریب خوب با $m+۱$ قدم



- مثال- دو خوشه
  - خوشه مقادیر ویژهٔ بزرگ دارای پنج عضو

سرعت همگرائی گرادیان مزدوج

تقریب بیشتر

$$\kappa(Q) = \|Q\|\|Q^{-۱}\| = \frac{\lambda_n}{\lambda_۱}$$

$$\|\boldsymbol{x}_i - \boldsymbol{x}^*\|_Q \leq ۲\left(\frac{\sqrt{\kappa(Q)} - ۱}{\sqrt{\kappa(Q)} + ۱}\right)^i \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_Q$$

مناسب در مواقع داشتن تقریب بالا و پایین مقدار ویژه‌ها

- مقایسه با گرادیان نزولی
- $\|\boldsymbol{x}_i - \boldsymbol{x}^*\|_Q = \left(\frac{\kappa(Q)-۱}{\kappa(Q)+۱}\right)^i \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_Q$

سرعت همگرائی گرادیان مزدوج

$$0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

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

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

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

- $\kappa(Q) = \|Q\| \left\|Q^{-۱}\right\|$
  - $\kappa(Q)$ کوچک: «خوش وضعیتی»
  - $\kappa(Q)$ بزرگ: بدوضعیتی»!
  - گاهی اوقات امکان تعریف $\kappa(Q)$ با مقادیر ویژه
    - چه مواقعی؟

سرعت همگرائی گرادیان مزدوج

$$\|\boldsymbol{x}_{i+1} - \boldsymbol{x}^*\| \leq \frac{\sqrt{\lambda_n} - \sqrt{\lambda_1}}{\sqrt{\lambda_n} + \sqrt{\lambda_1}} \|\boldsymbol{x}_i - \boldsymbol{x}^*\|$$

در صورت غلبهٔ λ_n بر λ_1

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

تابع درجه دو
- بیضی‌های متحدالمرکز

هر چه $\kappa(Q)$ بزرگتر
- بیضی کشیده‌تر
- کاهش همگرائی

منابع

[نازه‌دل]

[لوئنبرگر]