ارتقا روش Barzilai-Borwein با ویژگی خاتمه در متناهی قدم برای مسائل درجه ۲ در دو بعد

روش BB با قابلیت خاتمه متناهی قدمی

کسری خوشجو

دانشگاه صنعتی شریف

۶ خرداد ۱۴۰۲

روش Barzilai-Borwein

- دو روش انتخاب طول قدم :

$$\alpha_k^{BB1} = \operatorname*{argmin}_{\alpha} \|\alpha^{-1} s_{k-1} - y_{k-1}\| = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}$$

$$\alpha_k^{BB2} = \operatorname*{argmin}_{\alpha} \|s_{k-1} - \alpha y_{k-1}\| = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}$$

که در آن $s_{k-1} = x_k - x_{k-1}, y_{k-1} = g_k - g_{k-1}$

خاتمه متناهی قدمی

تعریف (خاتمه متناهی قدمی دو بعدی)

می‌گوییم یک روش انتخاب طول گام دارای خاصیت "خاتمه متناهی قدمی در مسائل درجه ۲ دو بعدی" دارد اگر بتوان تضمین کرد در تعدادی متناهی قدم از الگوریتم گرادیان کاهشی به نقطه بهینه برسد.

آزمون روی مساله درجه ۲

مساله زیر را درنظر بگیرید:

$$g(x) = \frac{۱}{۲} x^T A x - b^T x,$$
$$A \in S^n_{++}$$

آزمون روی مساله درجه ۲

اگر بدانیم:

$$Ag_{k+1} = \lambda g_{k+1}$$

و همچنین فرض کنیم

$$\alpha_{k+2} = \frac{(A^{\mu}g_{k+i})^T(A^{\mu}g_{k+i})}{(A^{\mu}g_{k+i})^T A(A^{\mu}g_{k+i})}$$

در اینصورت می‌توان گفت:

$$\exists i \in \{1, 2, 3\} : g_{k+i} = 0$$

آزمون روی مساله درجه ۲

توابع حقیقی $\psi_1, \psi_2, \psi_3, \psi_4$ را طوری درنظر بگیریم که:

$$\psi_1(x)\psi_2(x) = \psi_3(x)\psi_4(x)$$

دراینصورت خواهیم داشت

$$g_{\nu_1(k)}^T \psi_1(A) g_{k+1} \cdot g_{\nu_2(k)}^T \psi_2(A) g_{k+1} = g_{\nu_1(k)}^T \psi_3(A) g_{k+1} \cdot g_{\nu_2(k)}^T \psi_4(A) g_{k+1}$$

به طوری که $\nu_i(s) \in \{1, 2, \ldots, s\}$

آزمون روی مساله درجه ۲

از معادله $g_{k+1} = (I - \alpha_k A) g_k$ نتیجه می‌گیریم:

$$g^T_{\nu_1(k)} \psi_1(A) g_{k+1} \cdot g^T_{\nu_2(k)} \psi_2(A) g_{k+1} = g^T_{\nu_1(k)} \psi_3(A) g_{k+1} \cdot g^T_{\nu_2(k)} \psi_4(A) g_{k+1}$$

$$\Leftrightarrow$$

$$g^T_{\nu_1(k)} \psi_1(A)(I - \alpha_k A) g_k \cdot g^T_{\nu_2(k)} \psi_2(A)(I - \alpha_k A) g_k$$
$$= g^T_{\nu_1(k)} \psi_3(A)(I - \alpha_k A) g_k \cdot g^T_{\nu_2(k)} \psi_4(A)(I - \alpha_k A) g_k$$

آزمون روی مساله درجه ۲

بازنویسی برحست α_k:

$$\phi_۱ \alpha_k^۲ - \phi_۲ \alpha_k + \phi_۳ = ۰$$

به طوری که

$$\begin{aligned}
\phi_1 &= g_{\nu_1(k)}^{\mathsf{T}} \psi_1(A) A g_k \cdot g_{\nu_2(k)}^{\mathsf{T}} \psi_2(A) A g_k - g_{\nu_1(k)}^{\mathsf{T}} \psi_3(A) A g_k \cdot g_{\nu_2(k)}^{\mathsf{T}} \psi_4(A) A g_k, \\
\phi_2 &= g_{\nu_1(k)}^{\mathsf{T}} \psi_1(A) g_k \cdot g_{\nu_2(k)}^{\mathsf{T}} \psi_2(A) A g_k + g_{\nu_1(k)}^{\mathsf{T}} \psi_1(A) A g_k \cdot g_{\nu_2(k)}^{\mathsf{T}} \psi_2(A) g_k \\
&\quad - g_{\nu_1(k)}^{\mathsf{T}} \psi_3(A) g_k \cdot g_{\nu_2(k)}^{\mathsf{T}} \psi_4(A) A g_k - g_{\nu_1(k)}^{\mathsf{T}} \psi_3(A) A g_k \cdot g_{\nu_2(k)}^{\mathsf{T}} \psi_4(A) g_k, \\
\phi_3 &= g_{\nu_1(k)}^{\mathsf{T}} \psi_1(A) g_k \cdot g_{\nu_2(k)}^{\mathsf{T}} \psi_2(A) g_k - g_{\nu_1(k)}^{\mathsf{T}} \psi_3(A) g_k \cdot g_{\nu_2(k)}^{\mathsf{T}} \psi_4(A) g_k.
\end{aligned}$$

که نتیجه می‌دهد

$$\alpha_k = \frac{۲}{\frac{\phi_۲}{\phi_۳} \pm \sqrt{\left(\frac{\phi_۲}{\phi_۳}\right)^۲ - ۴\frac{\phi_۱}{\phi_۳}}}$$

آزمون روی مساله درجه ۲

قضیه ((۱.۲)شرط کافی برای داشتن خاصیت خاتمه متناهی قدمی)

در مساله دو بعدی درجه ۲ اگر قدم k قید ۱ را برقرار کند به طوری که $\psi_1(z)\psi_2(z) = \psi_3(z)\psi_4(z)$ و اگر α_{k+2} از رابطه ۲ بدست آید، می‌توانیم نتیجه بگیریم $g_{k+i} = 0$ برای $1 \leq i \leq 3$

$$\begin{aligned} & g^T_{\nu_1(k)}\psi_1(A)(I-\alpha_k A)g_k \cdot g^T_{\nu_2(k)}\psi_2(A)(I-\alpha_k A)g_k \\ &= g^T_{\nu_1(k)}\psi_3(A)(I-\alpha_k A)g_k \cdot g^T_{\nu_2(k)}\psi_4(A)(I-\alpha_k A)g_k \end{aligned} \tag{1}$$

$$\alpha_{k+2} = \frac{(A^\mu g_{k+i})^T(A^\mu g_{k+i})}{(A^\mu g_{k+i})^T A(A^\mu g_{k+i})} \tag{2}$$

روش های دارای ویژگی خاتمه متناهی قدمی

از قضیه (2.1) می توانیم ببینیم:

- روش "Steepest Descent" یا "عمیق ترین کاهش" و روش "BB" دارای خاصیت خاتمه متناهی قدمی هستند چرا که در معادله2 صدق می کنند.

تعریف طول قدم بهبود یافته

فرض کنیم طول قدم ما برابر باشد با α_k^{new}:

$$\alpha_k^{\text{new}} = \frac{۲}{\frac{\phi_۲}{\phi_۳} + \sqrt{\left(\frac{\phi_۲}{\phi_۳}\right)^۲ - ۴\frac{\phi_۱}{\phi_۳}}}$$

$$\text{where}: \frac{\phi_۱}{\phi_۳} = \frac{\alpha_{k-۱}^{BB۲} - \alpha_k^{BB۲}}{\alpha_{k-۱}^{BB۲}\alpha_k^{BB۲}(\alpha_{k-۱}^{BB۱} - \alpha_k^{BB۱})}$$

$$\text{and}\,\frac{\phi_۲}{\phi_۳} = \frac{\alpha_{k-۱}^{BB۱}\alpha_{k-۱}^{BB۲} - \alpha_k^{BB۲}\alpha_k^{BB۱}}{\alpha_{k-۱}^{BB۲}\alpha_k^{BB۲}(\alpha_{k-۱}^{BB۱} - \alpha_k^{BB۱})}$$

- برای برخی $\psi_۱, \psi_۲, \psi_۳, \psi_۴$ و $\nu_۱, \nu_۲$ α_k^{new} می‌توان نشان داد که شرایط قضیه(۱.۲) برای این طول قدم برقرار است.
- α_k^{new} خوش تعریف است.

α_k^{new} خوش تعریف است

قضیه (2.9) (کران های α_k^{new})

مقدار α_k^{new} همواره تعریف شده است و به طور خاص داریم:

وقتی $\phi_۲ \neq ۰$, $\phi_۱/\phi_۳ \geq ۰$:

$$\phi_۳/\phi_۲ \leq \alpha_k^{new} \leq \min\{\alpha_k^{BB۲}, \alpha_{k-۱}^{BB۲}\}$$

و وقتی $\phi_۲ \neq ۰$, $\phi_۱/\phi_۳ < ۰$:

$$\max\{\alpha_k^{BB۲}, \alpha_{k-۱}^{BB۲}\} \leq \alpha_k^{new} \leq |\phi_۳/\phi_۱|$$

حل مسائل بدون قید

- در مطالعات بسیاری دیده شده که انتخاب قدم به قدم $\alpha_k^{BB۱}, \alpha_k^{BB۲}$ در آزمایش های عددی بهتر از صرف استفاده از یکی از طول قدم هاست
- طول قدم خود را با دانستن این موضوع بهبود می‌بخشیم:

$$\alpha_k = \begin{cases} \min\{\alpha_{k-۱}^{BB۲}, \alpha_k^{BB۲}, \alpha_k^{\text{new}}\} & \text{if } \frac{\alpha_k^{BB۲}}{\alpha_k^{BB۱}} < \tau_k \\ \alpha_k^{BB۱} & \text{otherwise} \end{cases}$$

با تعریف مناسبی از دنباله کران های $\{\tau_k\}_k$
(Zhue et al. 2006 (Gradient Methods with Adaptive Step-Sizes))

- از روش جستجوی خطی غیر یکنوای GLL برای کنترل نرخ حرکت در جهت کاهشی بهره می‌بریم.

$$f(x_k + \lambda_k d_k) \le f_r + \sigma \lambda_k g_k^T d_k$$
$$f_r = max_{k-M\le i\le k} f(x_k)$$

الگوریتم (3.1)

Algorithm 3.1 A gradient method for unconstrained optimization.

Initialization: $x_1 \in \mathbb{R}^n$, $\alpha_{\max} \geq \alpha_{\min} > 0$, $\alpha_1 \in [\alpha_{\min}, \alpha_{\max}]$, $\tau_1 > 0$, $\gamma > 1$, $\epsilon, \sigma, \delta \in (0,1)$, $M \in \mathbb{N}$, $k := 1$.

while $\|g_k\|_\infty > \epsilon$ **do**
  $d_k = -g_k$, $\lambda_k = \alpha_k$
  $f_r = \max_{0 \leq i \leq \min\{k, M-1\}} f(x_{k-i})$
  while the condition (3.6) does not hold **do**
    $\lambda_k = \delta \lambda_k$
  end while
  $x_{k+1} = x_k + \lambda_k d_k$
  if $s_k^\mathsf{T} y_k > 0$ **then**
    if $\boxed{\alpha_k^{BB2}/\alpha_k^{BB1} < \tau_k}$ and $s_{k-1}^\mathsf{T} y_{k-1} > 0$ **then**
      if $\alpha_{k+1}^{new} > 0$ **then**
        $\alpha_{k+1} = \min\{\alpha_k^{BB2}, \alpha_{k+1}^{BB2}, \alpha_{k+1}^{new}\}$
      else
        $\alpha_{k+1} = \min\{\alpha_k^{BB2}, \alpha_{k+1}^{BB2}\}$
      end if
      $\boxed{\tau_{k+1} = \tau_k/\gamma}$
    else
      $\alpha_{k+1} = \alpha_{k+1}^{BB1}$
      $\boxed{\tau_{k+1} = \tau_k \gamma}$
    end if
  else
    $\alpha_{k+1} = \min\{1/\|g_k\|_\infty, \|x_k\|_\infty/\|g_k\|_\infty\}$
  end if
  Chop extreme values of the stepsize such that $\alpha_{k+1} \in [\alpha_{\min}, \alpha_{\max}]$
  $k = k + 1$
end while

تحلیل نرخ همگرایی

- همگرایی کلی (global convergence):
  از قضیه (2.2) در
  Birgin, et al(2000), Nonmonotone spectral projected gradient methods on convex sets
- همگرایی R-linear برای توابع قویا محدب:
  از قضیه(3.4) در
  Huang and H. Liu(2015), On the rate of convergence of projected Barzilai–Borwein methods

$$|x_k - L| \le \varepsilon_k \qquad \lim_{k\to\infty} \frac{|x_{k+1} - L|}{|x_k - L|^q} = \mu$$

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \ge m\|x - y\|^2$$

مقادیر ویژه کوچک

$$\min_{X\in\mathbb{R}^{n\times r}} tr\{X^TAX(X^TX)^{-۱}\}$$

$$\sim$$

$$\min_{X\in\mathbb{R}^{n\times r}} \frac{۱}{۴}tr\{X^TXX^TX\} + \frac{۱}{۲}tr\{X^T(A-\mu I)X\}$$

مقادیر ویژه کوچک

- تعریف طول قدم در مساله ماتریسی

$$\alpha_k^{MBB1} = \frac{\mathrm{tr}(S_{k-1}^{\mathsf{T}} S_{k-1})}{\mathrm{tr}(S_{k-1}^{\mathsf{T}} Y_{k-1})} \qquad \alpha_k^{MBB2} = \frac{\mathrm{tr}(S_{k-1}^{\mathsf{T}} Y_{k-1})}{\mathrm{tr}(Y_{k-1}^{\mathsf{T}} Y_{k-1})},$$

- روش جستجوی خطی غیر یکنوای Dai-Flecher برای کاهش تعداد مراتب محاسبه f

الگوریتم (4.1)

Algorithm 4.1 A gradient method for extreme eigenvalue problems.

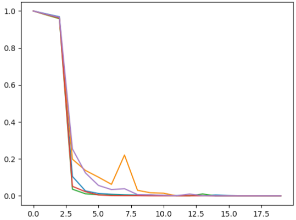
Initialization: $X_1 \in \mathbb{R}^{n\times r}$, $\tau_1 > 0$, $\alpha_{\max} \geq \alpha_{\min} > 0$, $\alpha_1 \in [\alpha_{\min}, \alpha_{\max}]$, $\tau_1 > 0$, $\gamma > 1$, $\epsilon, \sigma, \delta \in (0,1)$, $M \in \mathbb{N}$, $m = 0$, $f_r = f_{best} = f_c = P_\mu(X_1)$, $k := 1$.

while $\|\nabla P_\mu(X_k)\|_F > \epsilon$ **do**
  if $P_\mu(X_k) < f_{best}$ **then**
    $f_{best} = P_\mu(X_k)$, $f_c = P_\mu(X_k)$, $m = 0$
  else
    $f_c = \max\{f_c, P_\mu(X_k)\}$, $m = m + 1$
    if $m = M$ **then**
      $f_r = P_\mu(X_k)$, $f_c = P_\mu(X_k)$, $m = 0$
    end if
  end if
  $d_k = -\nabla P_\mu(X_k)$, $\lambda_k = \alpha_k$
  while the condition (3.6) does not hold **do**
    $\lambda_k = \delta\lambda_k$
  end while
  $X_{k+1} = X_k + \lambda_k d_k$
  if $\mathrm{tr}(S_k^\mathsf{T} Y_k) > 0$ **then**
    if $\alpha_k^{MBB2}/\alpha_k^{MBB1} < \tau_k$ and $\mathrm{tr}(S_{k-1}^\mathsf{T} Y_{k-1}) > 0$ **then**
      if $\alpha_{k+1}^{Mnew} > 0$ **then**
        $\alpha_{k+1} = \min\{\alpha_k^{MBB2}, \alpha_{k+1}^{MBB2}, \alpha_{k+1}^{Mnew}\}$
      else
        $\alpha_{k+1} = \min\{\alpha_k^{MBB2}, \alpha_{k+1}^{MBB2}\}$
      end if
      $\tau_{k+1} = \tau_k/\gamma$
    else
      $\alpha_{k+1} = \alpha_{k+1}^{MBB1}$
      $\tau_{k+1} = \tau_k\gamma$
    end if
  else
    $\alpha_{k+1} = |\alpha_{k+1}^{MBB1}|$
  end if
  Chop extreme values of the stepsize such that $\alpha_{k+1} \in [\alpha_{\min}, \alpha_{\max}]$
  $k = k + 1$
end while

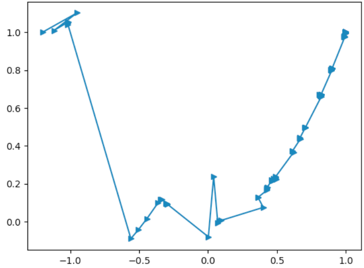
مسائل مقید خاص

- مقید به مکعب $l_n \leq x \leq u_n$
- مقید به مکعب و صفحه آفین $l_n \leq x \leq u_n, Ax = b$

https://github.com/KKasra/Enhanced-BB-method



مسائل درجه دوم ۱۰۰ بعدی تصادفی



rosenbrock badly scaled problem