



بهینه سازی
مبانی بهینه سازی نامقید
گرا دیان مزدوج

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

گرادیان مزدوج

هستنز و استیفل ۱۹۵۱

لنکوزس ۱۹۵۲

بهبود گرادیان نزولی با جلوگیری از تکرار مراحل

استفاده از تاریخچه قدم قبلی جهت سریع کردن همگرایی

مشابه شدیدترین نزول به جز در محاسبه کردن قدمها

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

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

حرکت در جهات متعامد
▪ جهت مزدوج با مسیر قبلی

متعامد یا مزدوجی دو بردار

$$\mathbf{p}_i^T Q \mathbf{p}_j = 0$$

جلوگیری از افزونگی

گرادیان مزدوج

هستنز و استیفل ۱۹۵۱

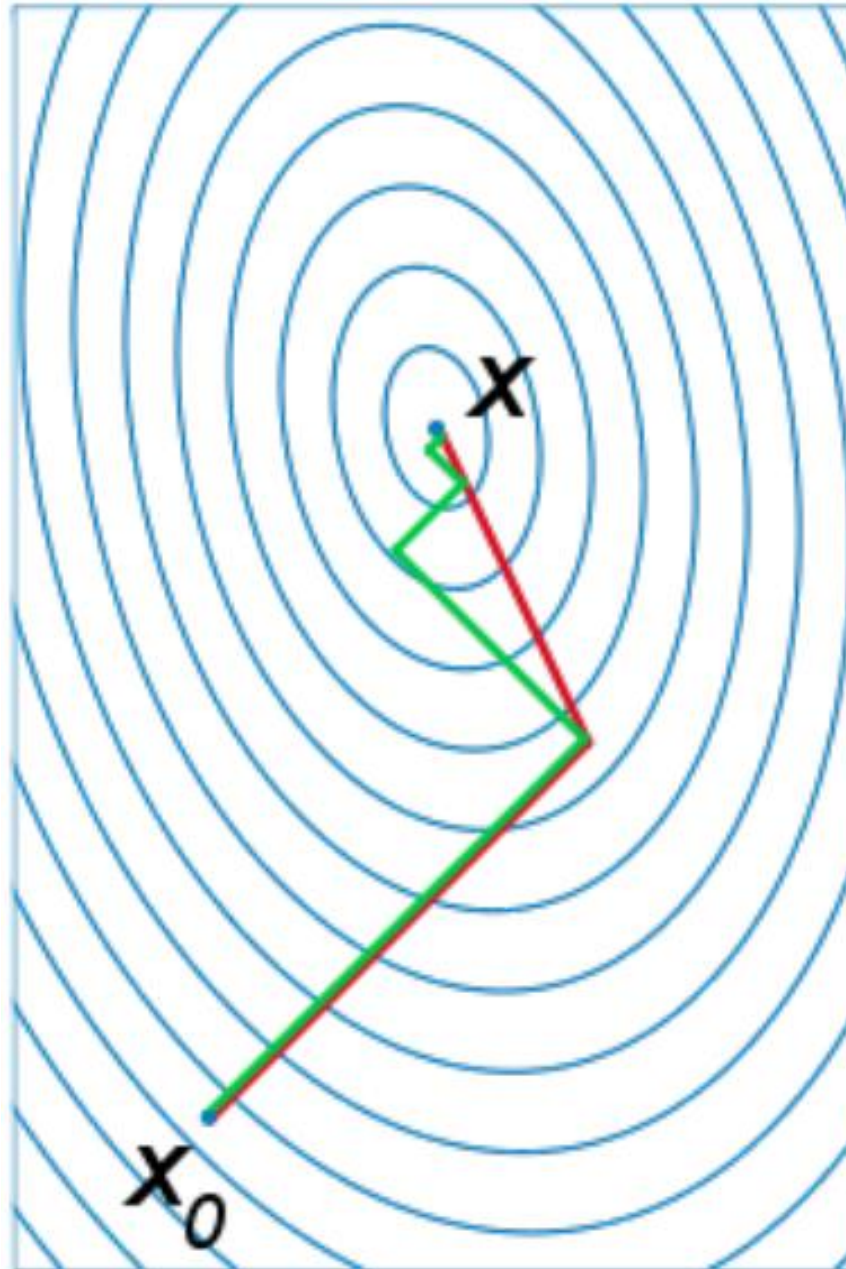
لنکوزس ۱۹۵۲

بهبود گرادیان نزولی با جلوگیری از تکرار مراحل
استفاده از تاریخچه قدم قبلی جهت سریع کردن
مشابه شدیدترین نزول به جز در محاسبه کردن

حرکت در جهات متعامد
جهت مزدوج با مسیر قبلی

متعامد یا مزدوجی دو بردار

جلوگیری از افزونگی



ضرب داخلی

$$\mathbf{a}, \mathbf{b} \in \mathbb{R}^n \Rightarrow \mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b} = \sum a_i b_i$$

$$f, g \Rightarrow f \cdot g = \int_T f(x) \bar{g}(x) dx$$

$$\mathbf{x}, \mathbf{y} \in \mathbb{R}^n, Q \text{ ماتریس مثبت معین} \Rightarrow \\ \Rightarrow \mathbf{x} \cdot \mathbf{y}_Q = \mathbf{x}^T Q \mathbf{y}$$

مسیرهای مزدوج

تعریف

Q ماتریس مثبت معین است. بردارهای $\{p_1, p_2, \dots, p_k\}$ با شرط $p_i \in \mathbb{R}^n, p_i \neq 0$ متعامد- Q (مزدوج- Q) نسبت به Q هستند اگر

$$p_i^T Q p_j = 0, \forall i \neq j$$

$Q=0$

$Q=I$

لم

Q مثبت معین باشد. اگر $\{p_1, p_2, \dots, p_k\}$ متعامد- Q باشند، آن گاه همگی مستقل خطی از یکدیگرند.

اثبات

اهمیت مزدوج-Q

هدف

$$\frac{1}{2} \mathbf{x}^T Q \mathbf{x} - \mathbf{b}^T \mathbf{x}$$

فرض $Q \in \mathbb{R}^{n \times n}$ مثبت معین

پاسخ به مسئله بالا، پاسخ به معادله خطی روبرو نیز است: $Q\mathbf{x} = \mathbf{b}, \mathbf{x} \in \mathbb{R}^n$

فرض \mathbf{x}^* پاسخ مسئله

$\{\mathbf{p}_0, \mathbf{p}_2, \dots, \mathbf{p}_{n-1}\}$ متعامد-Q.

چون $\{\mathbf{p}_0, \mathbf{p}_2, \dots, \mathbf{p}_{n-1}\}$ مستقل هستند

$$\mathbf{x}^* = \alpha_0 \mathbf{p}_0 + \dots + \alpha_{n-1} \mathbf{p}_{n-1}$$

اهمیت مزدوج-Q

$$\mathbf{x}^* = \alpha_0 \mathbf{p}_0 + \cdots + \alpha_{n-1} \mathbf{p}_{n-1}$$

پس

$$\mathbf{p}_i^T Q \mathbf{x}^* = \mathbf{d}_i^T Q (\alpha_0 \mathbf{p}_0 + \cdots + \alpha_{n-1} \mathbf{p}_{n-1}) = \alpha_i \mathbf{p}_i^T Q \mathbf{p}_i$$

$$\alpha_i = \frac{\mathbf{p}_i^T Q \mathbf{x}^*}{\mathbf{p}_i^T Q \mathbf{p}_i} = \frac{\mathbf{p}_i^T \mathbf{b}}{\mathbf{p}_i^T Q \mathbf{p}_i}$$

عدم نیاز به دانستن \mathbf{x}^* برای یافتن α_i

متعامد بودن معیار کافی نیست

▪ نیاز به متعامد-Q

اهمیت مزدوج-Q

$$\mathbf{x}^* = \alpha_0 \mathbf{p}_0 + \cdots + \alpha_{n-1} \mathbf{p}_{n-1}$$

$$\alpha_i = \frac{\mathbf{p}_i^T Q \mathbf{x}^*}{\mathbf{p}_i^T Q \mathbf{p}_i} = \frac{\mathbf{p}_i^T \mathbf{b}}{\mathbf{p}_i^T Q \mathbf{p}_i}$$
$$\mathbf{x}^* = \sum_{i=0}^{n-1} \alpha_i \mathbf{p}_i = \sum_{i=0}^{n-1} \frac{\mathbf{p}_i^T \mathbf{b}}{\mathbf{p}_i^T Q \mathbf{p}_i} \mathbf{p}_i$$

صرفاً احتیاج به ضرب داخلی

- امکان یافتن نقطه کمینه‌سازی با استفاده از روش تکرار مراحل و عددی در n مرحله
- در مرحله i -ام مقدار $\alpha_i \mathbf{p}_i$ افزوده می‌شود
 - تعمیم بیشتر با آغاز از نقطه دلخواه \mathbf{x}_0

کمینه‌سازی توابع درجه دو

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

فرض بر داشتن $\{\mathbf{p}_i\}_{i=1}^n$ برابر با Q -مزدوج. فرض $\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{p}_i$

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|_Q^2 - \mathbf{b}^T \mathbf{x} = \sum_{i=1}^n \underbrace{\left(\frac{1}{2} \alpha_i^2 \|\mathbf{p}_i\|_Q^2 - \alpha_i \mathbf{b}^T \mathbf{p}_i \right)}_{\phi_i(\alpha_i)}$$

$$f(\mathbf{x}(\boldsymbol{\alpha})) = \sum_{i=1}^n \phi_i(\alpha_i)$$

جداپذیر بر الف

کمینه‌سازی توابع درجه دو

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

فرض بر داشتن $\{\mathbf{p}_i\}_{i=1}^n$ برابر با Q -مزدوج. فرض $\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{p}_i$

$$f(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}\|_Q^2 - \mathbf{b}^T \mathbf{x} = \sum_{i=1}^n \underbrace{\left(\frac{1}{2} \alpha_i^2 \|\mathbf{p}_i\|_Q^2 - \alpha_i \mathbf{b}^T \mathbf{p}_i \right)}_{\phi_i(\alpha_i)}$$

$$\min_{\alpha} f(\mathbf{x}(\alpha)) = \sum_{i=1}^n \min_{\alpha_i} \phi_i(\alpha_i)$$

n دنباله از کمینه‌سازی‌های جستجو خط یک-بعدی

روش جهت مزدوج

$$\min_{\alpha} f(x(\alpha)) = \sum_{i=1}^n \min_{\alpha_i} \phi_i(\alpha_i)$$

n دنباله از کمینه‌سازی‌های جستجو خط یک-بعدی

بسط خاصیت زیرفضا- d /امه

$$\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{p}_i$$
$$\mathbf{x} = \mathbf{x}_0 + \sum_{i=1}^k \alpha_i \mathbf{p}_i$$
$$G_k = \left\{ \mathbf{x} = \mathbf{x}_0 + \sum_{i=1}^k \alpha_i \mathbf{p}_i, \forall \alpha \right\}$$

پس از k مرحله از روش مزدوج

$$\mathbf{x}_k = \operatorname{argmin}_{\mathbf{x} \in G_k} f(\mathbf{x})$$

بسط خاصیت زیرفضا-ا/د/امه

$$\mathbf{x} = \mathbf{x}_0 + \sum_{i=1}^k \alpha_i \mathbf{p}_i$$

$$\boldsymbol{\alpha} \in \mathbb{R}^k, P = (\mathbf{p}_1 \mathbf{p}_2 \dots \mathbf{p}_k) \Rightarrow \mathbf{x} = \mathbf{x}_0 + P\boldsymbol{\alpha}$$

$$\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} f(\mathbf{x}_0 + P\boldsymbol{\alpha})$$

$$\mathbf{x}^* = \mathbf{x}_0 + P\boldsymbol{\alpha}^*$$

$$\nabla_{\boldsymbol{\alpha}} = P^T \nabla_{\mathbf{x}} f(\underbrace{\mathbf{x}_0 + P\boldsymbol{\alpha}}_{\mathbf{x}})$$

$$= P^T \nabla_{\mathbf{x}} f(\mathbf{x}^*) \stackrel{\mathbf{x}}{=} 0$$

بسط خاصیت زیرفضا- d/d امه

$$\mathbf{x} = \mathbf{x}_0 + \sum_{i=1}^k \alpha_i \mathbf{p}_i$$

$$\alpha \in \mathbb{R}^k, P = (\mathbf{p}_1 \mathbf{p}_2 \dots \mathbf{p}_k) \Rightarrow \mathbf{x} = \mathbf{x}_0 + P\alpha$$

$$\alpha^* = \underset{\alpha}{\operatorname{argmin}} f(\mathbf{x}_0 + P\alpha)$$

$$\mathbf{x}^* = \mathbf{x}_0 + P\alpha^*$$

$$\nabla_{\alpha} = P^T \nabla_{\mathbf{x}} f(\mathbf{x}_0 + P\alpha)$$

$$= P^T \nabla_{\mathbf{x}} f(\mathbf{x}^*) = 0 \Rightarrow \begin{cases} \mathbf{p}_1^T \nabla_{\mathbf{x}} f(\mathbf{x}^*) = 0 \\ \vdots \\ \mathbf{p}_k^T \nabla_{\mathbf{x}} f(\mathbf{x}^*) = 0 \end{cases}$$

بسط خاصیت زیرفضا- d/d امه

$$\nabla_{\alpha} = P^T \nabla_x f(\underbrace{\mathbf{x}_0 + P\alpha}_x)$$

$$= P^T \nabla_x f(\mathbf{x}^*) = 0 \Rightarrow \begin{cases} \mathbf{p}_1^T \nabla_x f(\mathbf{x}^*) = 0 \\ \vdots \\ \mathbf{p}_k^T \nabla_x f(\mathbf{x}^*) = 0 \end{cases}$$

گرادیان جدید متعامد بر تمامی جهتهای قبلی \mathbf{p}_i و $i=1, \dots, k$

کمینه‌سازی توابع درجه دو

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

آغاز:

$$\mathbf{p}_0 = -\mathbf{g}_0 = -Q \mathbf{x}_0 + \mathbf{b}$$

مرحله k -ام

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$$

نیاز به جستجو خط کامل

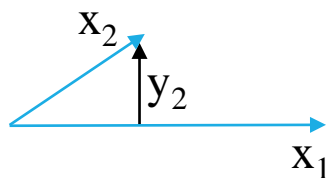
$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \sum_{j=0}^k \left(\frac{\mathbf{g}_{k+1}^T Q \mathbf{p}_j}{\mathbf{p}_j^T Q \mathbf{p}_j} \mathbf{p}_j \right)$$

استفاده از گرام-اشمیت برای متعامد- Q مربوط به $\{\mathbf{p}_j\}_{j=0}^k$

متعامدسازی گرام-اشمیت - ادامه

{ x_1, x_2, \dots, x_n } بردارهای مستقل خطی

ایجاد محورهای (پایه‌های) متعامد



$$y_1 = x_1$$

$$y_2 = x_2 - \left(x_2 \cdot \frac{y_1}{\|y_1\|}\right) \frac{y_1}{\|y_1\|} = x_2 - \frac{x_2 \cdot y_1}{\|y_1\|^2} y_1$$

⋮

$$y_j = x_j - \sum_{i=1}^{j-1} \frac{x_j \cdot y_i}{\|y_i\|^2} y_i$$

کمینه‌سازی توابع درجه دو

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \sum_{j=0}^k \left(\frac{\mathbf{g}_{k+1}^T Q \mathbf{p}_j}{\mathbf{p}_j^T Q \mathbf{p}_j} \mathbf{p}_j \right)$$

ساده‌سازی ۱

$$\mathbf{x}_{j+1} = \mathbf{x}_j + \alpha_j \mathbf{p}_j \Rightarrow$$

$$\mathbf{p}_j = \frac{1}{\alpha_j} (\mathbf{x}_{j+1} - \mathbf{x}_j)$$

کمینه‌سازی توابع درجه دو

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \sum_{j=0}^k \left(\frac{\mathbf{g}_{k+1}^T Q \mathbf{p}_j}{\mathbf{p}_j^T Q \mathbf{p}_j} \mathbf{p}_j \right)$$

ساده‌سازی ۱

$$\mathbf{p}_j = \frac{1}{\alpha_j} (\mathbf{x}_{j+1} - \mathbf{x}_j) \Rightarrow Q \mathbf{p}_j = \frac{1}{\alpha_j} Q (\mathbf{x}_{j+1} - \mathbf{x}_j)$$

کمینه‌سازی توابع درجه دو

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \sum_{j=0}^k \left(\frac{\mathbf{g}_{k+1}^T Q \mathbf{p}_j}{\mathbf{p}_j^T Q \mathbf{p}_j} \mathbf{p}_j \right)$$

ساده‌سازی ۱

$$\mathbf{p}_j = \frac{1}{\alpha_j} (\mathbf{x}_{j+1} - \mathbf{x}_j) \Rightarrow Q \mathbf{p}_j = \frac{1}{\alpha_j} Q (\mathbf{x}_{j+1} - \mathbf{x}_j) = \frac{1}{\alpha_j} (\mathbf{g}_{j+1} - \mathbf{g}_j)$$

\mathbf{g}_{k+1} عمود بر $\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_k \leftarrow$ پس آخرین بخش در جمع مقادیر بالای صفحه کافی است

کمینه‌سازی توابع درجه دو

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \sum_{j=0}^k \left(\frac{\mathbf{g}_{k+1}^T Q \mathbf{p}_j}{\mathbf{p}_j^T Q \mathbf{p}_j} \mathbf{p}_j \right)$$

ساده‌سازی ۱

(الف)

$$\mathbf{p}_j = \frac{1}{\alpha_j} (\mathbf{x}_{j+1} - \mathbf{x}_j) \Rightarrow Q \mathbf{p}_j = \frac{1}{\alpha_j} Q (\mathbf{x}_{j+1} - \mathbf{x}_j) = \frac{1}{\alpha_j} (\mathbf{g}_{j+1} - \mathbf{g}_j)$$

(ب) با قضیه بسط زیرفضا

کمینه‌سازی توابع درجه دو

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \sum_{j=0}^k \left(\frac{\mathbf{g}_{k+1}^T Q \mathbf{p}_j}{\mathbf{p}_j^T Q \mathbf{p}_j} \mathbf{p}_j \right)$$

ساده‌سازی ۱

$$\mathbf{p}_j = \frac{1}{\alpha_j} (\mathbf{x}_{j+1} - \mathbf{x}_j) \Rightarrow Q \mathbf{p}_j = \frac{1}{\alpha_j} Q (\mathbf{x}_{j+1} - \mathbf{x}_j) = \frac{1}{\alpha_j} (\mathbf{g}_{j+1} - \mathbf{g}_j)$$

(ب) با قضیه بسط زیرفضا

\mathbf{g}_{k+1} عمود بر $\mathbf{g}_0, \mathbf{g}_1, \dots, \mathbf{g}_k$ پس آخرین بخش کافی است

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \frac{\mathbf{g}_{k+1}^T (\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{p}_k^T (\mathbf{g}_{k+1} - \mathbf{g}_k)} \mathbf{p}_k$$

کمینه‌سازی توابع درجه دو

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \frac{\mathbf{g}_{k+1}^T(\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{p}_k^T(\mathbf{g}_{k+1} - \mathbf{g}_k)} \mathbf{p}_k$$

$$\beta_k = \frac{\mathbf{g}_{k+1}^T(\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{p}_k^T(\mathbf{g}_{k+1} - \mathbf{g}_k)}$$

ساده‌سازی ۲

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{p}_k$$

الف) \mathbf{p}_k عمود بر \mathbf{g}_{k+1} پس در مخرج داریم $\mathbf{p}_k^T(\mathbf{g}_{k+1} - \mathbf{g}_k) = \mathbf{p}_k^T(-\mathbf{g}_k)$

ب) همچنین $\mathbf{p}_k = -\mathbf{g}_k + \beta_{k-1} \mathbf{p}_{k-1}$ پس جایگذاری در معادله ساده شده مرحله قبل
 $(-\mathbf{g}_k + \beta_{k-1} \mathbf{p}_{k-1})^T(-\mathbf{g}_k) = \mathbf{g}_k^T \mathbf{g}_k$

$\beta_{k-1} \mathbf{p}_{k-1}$ عمود بر \mathbf{g}_k پس

$$\beta_k = \frac{\mathbf{g}_{k+1}^T(\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{g}_k^T \mathbf{g}_k}$$

کمینه‌سازی توابع درجه دو

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \frac{\mathbf{g}_{k+1}^T(\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{p}_k^T(\mathbf{g}_{k+1} - \mathbf{g}_k)} \mathbf{p}_k$$

$$\beta_k = \frac{\mathbf{g}_{k+1}^T(\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{p}_k^T(\mathbf{g}_{k+1} - \mathbf{g}_k)}$$

ساده‌سازی ۲

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{p}_k$$

الف) \mathbf{p}_k عمود بر \mathbf{g}_{k+1} پس در مخرج داریم $\mathbf{p}_k^T(\mathbf{g}_{k+1} - \mathbf{g}_k) = \mathbf{p}_k^T(-\mathbf{g}_k)$

ب) همچنین $\mathbf{p}_k = -\mathbf{g}_k + \beta_{k-1} \mathbf{p}_{k-1}$ پس جایگذاری در معادله ساده شده مرحله قبل
 $(-\mathbf{g}_k + \beta_{k-1} \mathbf{p}_{k-1})^T(-\mathbf{g}_k) = \mathbf{g}_k^T \mathbf{g}_k$

پس

$$\beta_k = \frac{\mathbf{g}_{k+1}^T(\mathbf{g}_{k+1} - \mathbf{g}_k)}{\mathbf{g}_k^T \mathbf{g}_k} = \frac{\mathbf{g}_{k+1}^T(\mathbf{g}_{k+1} - \mathbf{g}_k)}{\|\mathbf{g}_k\|_2^2} = \frac{\|\mathbf{g}_{k+1}\|_2^2}{\|\mathbf{g}_k\|_2^2}$$

کمینه‌سازی توابع درجه دو

$$\mathbf{p}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{p}_k$$

پس

$$\beta_k = \frac{\mathbf{g}_{k+1}^T (\mathbf{g}_{k+1} - \mathbf{g}_k)}{\|\mathbf{g}_k\|_2^2}$$

روش پولاک-ریبری

$$\beta_k = \frac{\|\mathbf{g}_{k+1}\|_2^2}{\|\mathbf{g}_k\|_2^2}$$

روش فلچر-ریوز

صورت پولاک دارای عملکرد بهتر در روش حل عددی توابع غیردرجه دو

الگوریتم گرادیان مزدوج

$$f: \mathbb{R}^n \rightarrow \mathbb{R}$$

▪ به دنبال یافتن کمینه

الگوریتم گرادیان مزدوج

مقداردهی اولیه \mathbf{x}_0 و $\mathbf{p}_0 = -\mathbf{g}_0 = \nabla f(\mathbf{x}_0)$

برای n بار

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

$$g(\alpha) = f(\mathbf{x}_i + \alpha_i \mathbf{p}_i) \quad \bullet$$

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

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

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

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

$\mathbf{x}_0 = \mathbf{x}_n$ و تکرار تا همگرا شدن

$\beta_k^{HS} = \frac{\mathbf{g}_{k+1}^\top \mathbf{y}_k}{\mathbf{d}_k^\top \mathbf{y}_k}$	(1952)	in the original (linear) CG paper of Hestenes and Stiefel [59]
$\beta_k^{FR} = \frac{\ \mathbf{g}_{k+1}\ ^2}{\ \mathbf{g}_k\ ^2}$	(1964)	first nonlinear CG method, proposed by Fletcher and Reeves [45]
$\beta_k^D = \frac{\mathbf{g}_{k+1}^\top \nabla^2 f(\mathbf{x}_k) \mathbf{d}_k}{\mathbf{d}_k^\top \nabla^2 f(\mathbf{x}_k) \mathbf{d}_k}$	(1967)	proposed by Daniel [39], requires evaluation of the Hessian $\nabla^2 f(\mathbf{x})$
$\beta_k^{PRP} = \frac{\mathbf{g}_{k+1}^\top \mathbf{y}_k}{\ \mathbf{g}_k\ ^2}$	(1969)	proposed by Polak and Ribière [84] and by Polyak [85]
$\beta_k^{CD} = \frac{\ \mathbf{g}_{k+1}\ ^2}{-\mathbf{d}_k^\top \mathbf{g}_k}$	(1987)	proposed by Fletcher [44], CD stands for “Conjugate Descent”
$\beta_k^{LS} = \frac{\mathbf{g}_{k+1}^\top \mathbf{y}_k}{-\mathbf{d}_k^\top \mathbf{g}_k}$	(1991)	proposed by Liu and Storey [67]
$\beta_k^{DY} = \frac{\ \mathbf{g}_{k+1}\ ^2}{\mathbf{d}_k^\top \mathbf{y}_k}$	(1999)	proposed by Dai and Yuan [27]
$\beta_k^N = \left(\mathbf{y}_k - 2\mathbf{d}_k \frac{\ \mathbf{y}_k\ ^2}{\mathbf{d}_k^\top \mathbf{y}_k} \right)^\top \frac{\mathbf{g}_{k+1}}{\mathbf{d}_k^\top \mathbf{y}_k}$	(2005)	proposed by Hager and Zhang [53]

TABLE 1.1
Various choices for the CG update parameter

روش مسیره‌های مزدوج

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

آغاز:

$$\mathbf{p}_0 = -\mathbf{g}_0 = -Q \mathbf{x}_0 + \mathbf{b}$$

مرحله i -ام

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

نیاز به جستجو خط کامل

$$\mathbf{p}_{i+1} = -\mathbf{g}_{i+1} + \sum_{j=0}^i \left(\frac{\mathbf{g}_{i+1}^T Q \mathbf{p}_j}{\mathbf{p}_j^T Q \mathbf{p}_j} \mathbf{p}_j \right)$$

استفاده از گرام-اشمیت برای متعامد- Q مربوط به $\{\mathbf{p}_j\}_{j=0}^i$

الگوریتم گرادیان مزدوج

$$f: \mathbb{R}^n \rightarrow \mathbb{R}$$

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

$$\mathbf{x}_i = \mathbf{x}_0 + \alpha_0 \mathbf{p}_0 + \cdots + \alpha_{i-1} \mathbf{p}_{i-1}$$

$$\mathbf{p}_i^T Q \mathbf{x}_i = \mathbf{p}_i^T Q \mathbf{x}_0 \implies \mathbf{p}_i^T Q (\mathbf{x}_i - \mathbf{x}_0) = 0$$

$$\mathbf{p}_i^T Q (\mathbf{x}^* - \mathbf{x}_0) = \mathbf{p}_i^T Q (\mathbf{x}^* - \mathbf{x}_i) = \mathbf{p}_i^T (\mathbf{b} - Q \mathbf{x}_i) = -\mathbf{p}_i^T \mathbf{g}_i$$

$$\mathbf{p}_i^T Q (\mathbf{x}^* - \mathbf{x}_0) = -\mathbf{p}_i^T \mathbf{g}_i$$

$$\mathbf{x}^* = \mathbf{x}_0 + \alpha_0 \mathbf{p}_0 + \cdots + \alpha_{n-1} \mathbf{p}_{n-1}$$

$$\mathbf{p}_i^T Q \mathbf{x}^* = \mathbf{p}_i^T Q \mathbf{x}_0 + \alpha_i \mathbf{p}_i^T Q \mathbf{p}_i$$

$$\mathbf{p}_i^T Q (\mathbf{x}^* - \mathbf{x}_0) = \alpha_i \mathbf{p}_i^T Q \mathbf{p}_i$$

$$\implies \alpha_i = -\frac{\mathbf{p}_i^T \mathbf{g}_i}{\mathbf{p}_i^T Q \mathbf{p}_i}$$

روش گرادینان مزدوج

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

آغاز:

$$\mathbf{p}_0 = -\mathbf{g}_0 = -Q \mathbf{x}_0 + \mathbf{b}$$

مرحله i-ام

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

روش گرادینان مزدوج

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

آغاز:

$$\mathbf{p}_0 = -\mathbf{g}_0 = -Q \mathbf{x}_0 + \mathbf{b}$$

مرحله i -ام

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

نیاز به جستجو خط کامل

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

روش گرادینان مزدوج

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

$$\mathbf{p}_i^T Q \mathbf{p}_{i+1} = \mathbf{p}_i^T Q (-\mathbf{g}_{i+1} + \beta_i \mathbf{p}_i)$$

$$0 = -\mathbf{p}_i^T Q \mathbf{g}_{i+1} + \beta_i \mathbf{p}_i^T Q \mathbf{p}_i$$

$$\mathbf{p}_i^T Q \mathbf{g}_{i+1} = \beta_i \mathbf{p}_i^T Q \mathbf{p}_i$$

$$\beta_i = \frac{\mathbf{p}_i^T Q \mathbf{g}_{i+1}}{\mathbf{p}_i^T Q \mathbf{p}_i} = \frac{\mathbf{g}_{i+1}^T Q \mathbf{p}_i}{\mathbf{p}_i^T Q \mathbf{p}_i}$$

الگوریتم گرادیان مزدوج

$$f: \mathbb{R}^n \rightarrow \mathbb{R}$$

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

▪ به دنبال یافتن کمینه

الگوریتم گرادیان مزدوج ۲

مقداردهی اولیه \mathbf{x}_0 و $\mathbf{g}_0 = Q\mathbf{x} - \mathbf{b}$ و $\mathbf{p}_0 = -\mathbf{g}_0$

برای n بار

$$\alpha_i = -\frac{\mathbf{g}_i^T \mathbf{p}_i}{\mathbf{p}_i^T Q \mathbf{p}_i} \bullet$$

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

$$\mathbf{g}_{i+1} = Q\mathbf{x}_{i+1} - \mathbf{b} \bullet$$

$$\beta_i = \frac{\mathbf{g}_{i+1}^T Q \mathbf{p}_i}{\mathbf{p}_i^T Q \mathbf{p}_i} \bullet$$

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

مثال

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{g}_0 = \nabla f(\mathbf{x}_0) = A\mathbf{x} - \mathbf{b} = -\mathbf{b} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$\mathbf{p}_0 = -\mathbf{g}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\alpha_0 = -\frac{\mathbf{g}_0^T \mathbf{p}_0}{\mathbf{p}_0^T Q \mathbf{p}_0} = -\frac{[-1 \ 0] \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{[1 \ 0] \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}} = \frac{1}{2}$$

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0 \mathbf{p}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}$$

$$\mathbf{g}_1 = Q\mathbf{x}_1 - \mathbf{b} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix}$$

$$\beta_0 = \frac{\mathbf{g}_1^T Q \mathbf{p}_0}{\mathbf{p}_0^T Q \mathbf{p}_0} = \frac{[0 \ -\frac{1}{2}] \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}{[1 \ 0] \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}} = \frac{1}{4}$$

$$\mathbf{p}_1 = -\mathbf{g}_1 + \beta_0 \mathbf{p}_0 = -\begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \end{bmatrix}$$

$$\alpha_1 = -\frac{\mathbf{g}_1^T \mathbf{p}_1}{\mathbf{p}_1^T Q \mathbf{p}_1} = -\frac{[0 \ -\frac{1}{2}] \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \end{bmatrix}}{[\frac{1}{4} \ \frac{1}{2}] \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \end{bmatrix}} = \frac{2}{3}$$

$$\mathbf{x}_2 = \mathbf{x}_1 + \alpha_1 \mathbf{p}_1 = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{3}{4} \\ \frac{1}{3} \end{bmatrix}$$

$$\mathbf{g}_2 = Q\mathbf{x}_2 - \mathbf{b} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} \frac{3}{4} \\ \frac{1}{3} \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

الگوریتم گرادیان مزدوج روش کاراتر

$$f: \mathbb{R}^n \rightarrow \mathbb{R}$$

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

▪ به دنبال یافتن کمینه

الگوریتم گرادیان مزدوج ۳

مقداردهی اولیه \mathbf{x}_0 و $\mathbf{g}_0 = Q\mathbf{x} - \mathbf{b}$ و $\mathbf{p}_0 = -\mathbf{g}_0$

تازمان $\mathbf{g}_i \neq 0$

$$\alpha_i = \frac{\mathbf{g}_i^T \mathbf{g}_i}{\mathbf{p}_i^T Q \mathbf{p}_i} \bullet$$

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

$$\mathbf{g}_{i+1} = \mathbf{g}_i + \alpha_i Q \mathbf{p}_i \bullet$$

$$\beta_i = \frac{\mathbf{g}_{i+1}^T \mathbf{g}_{i+1}}{r_i^T r_i} \bullet$$

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

مثال

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{g}_0 = \nabla f(\mathbf{x}_0) = A\mathbf{x} - \mathbf{b} = -\mathbf{b} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$\mathbf{p}_0 = -\mathbf{g}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\alpha_0 = \frac{\mathbf{g}_0^T \mathbf{g}_0}{\mathbf{p}_0^T Q \mathbf{p}_0} = \frac{\begin{bmatrix} -1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}} = \frac{1}{2}$$

$$\mathbf{x}_1 = \mathbf{x}_0 + \alpha_0 \mathbf{p}_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}$$

$$\mathbf{g}_1 = \mathbf{g}_0 + \alpha_0 Q \mathbf{p}_0 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix}$$

$$\beta_0 = \frac{\mathbf{g}_1^T \mathbf{g}_1}{\mathbf{g}_0^T \mathbf{g}_0} = \frac{\begin{bmatrix} 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix}}{\begin{bmatrix} -1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix}} = \frac{1}{4}$$

$$\mathbf{p}_1 = -\mathbf{g}_1 + \beta_0 \mathbf{p}_0 = \begin{bmatrix} 0 \\ \frac{1}{2} \\ 2 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ 2 \end{bmatrix}$$

$$\alpha_1 = \frac{\mathbf{g}_1^T \mathbf{g}_1}{\mathbf{p}_1^T Q \mathbf{p}_1} = \frac{2}{3}$$

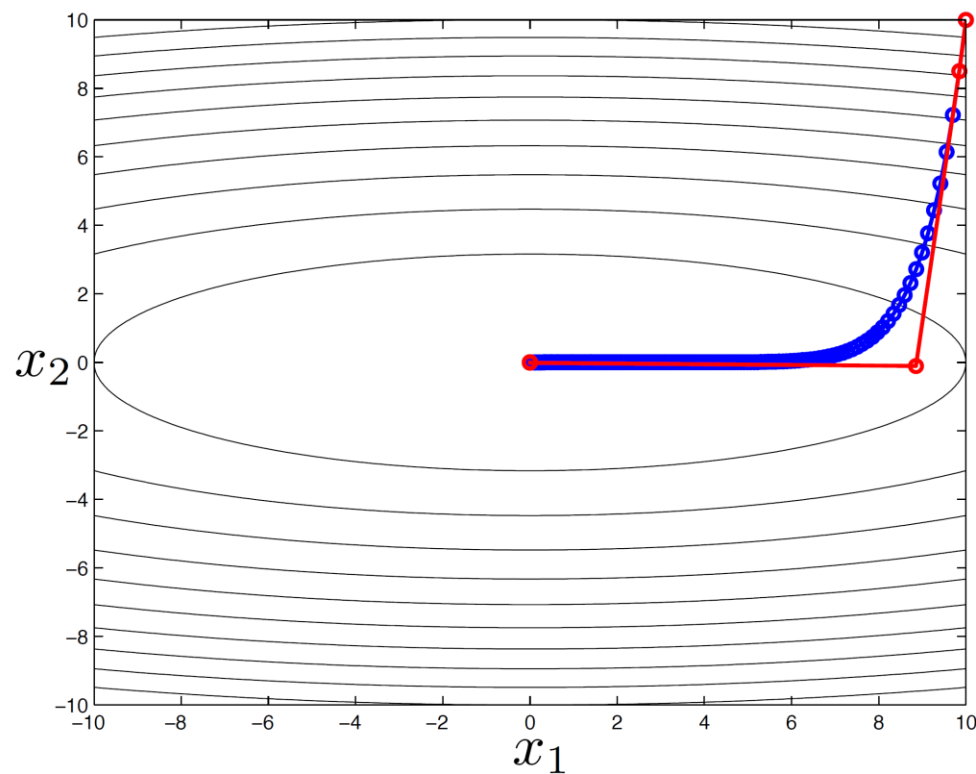
$$\mathbf{x}_2 = \mathbf{x}_1 + \alpha_1 \mathbf{p}_1 = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ 3 \end{bmatrix}$$

$$\mathbf{g}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

مثال - تندترین نزول در برابر گرادیان مزدوج

$$f(\mathbf{x}) = x_1^2 + 10x_2^2 \quad \mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix}, \mathbf{b} = 0$$

$$\alpha_i = 0.015$$



گرادیان مزدوج

«تصحیح» جهت شدیدترین نزول

$$p_i = -\nabla f(x_i) + \beta_i p_{i-1}$$

محاسبه β_i جهت مزدوج کردن p_i و p_{i-1}

- اجازه استفاده از جهت قبلی

مزایا

- کاراتر از گرادیان کاهشی
- پیاده‌سازی ساده به همان میزان
- صرفاً اطلاع مرتبه اول
- حافظه کم
- قابل استفاده جهت توابع غیرخطی

معایب

- سرعت همگرایی متوسط
- حساس به مقیاس گذاری

سرعت همگرایی گرادیان مزدوج

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

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_Q^2 \leq \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1} \right)^2 \|\mathbf{x}_0 - \mathbf{x}^*\|_Q^2$$

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

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

سرعت همگرایی گرادیان مزدوج

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

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

پیش شرطی

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

تغییر متغیرها

▪ فرض S معکوس پذیر و $\mathbf{y} = S\mathbf{x}$

جایگذاری در تابع هدف

$$\begin{aligned} \phi(\mathbf{y}) &= \frac{1}{2} \mathbf{y}^T (S^{-T} Q S^{-1}) \mathbf{y} - (S^{-T} \mathbf{b})^T \mathbf{y} \\ (S^{-T} Q S^{-1}) \mathbf{y} &= (S^{-T} \mathbf{b})^T \end{aligned}$$

سرعت همگرایی به جای Q وابسته به مقدار ویژه‌های ماتریس $S^{-T} Q S^{-1}$

▪ پس به دنبال S ی که مقدار ویژه‌های مناسبتر جهت همگرایی

▪ تبدیل به ماتریسی با عدد شرط کمتر

الگوریتم گرادیان مزدوج پیش شرطی

$$f: \mathbb{R}^n \rightarrow \mathbb{R}$$

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

الگوریتم گرادیان مزدوج پیش شرطی

M پیش شرط ساز و

مقداردهی اولیه \mathbf{x}_0 و $\mathbf{g}_0 = Q\mathbf{x}_0 - \mathbf{b}$

حل $M\mathbf{y}_0 = \mathbf{g}_0$

قرار دادن $\mathbf{p}_0 = -\mathbf{y}_0$

تازمان $\mathbf{g}_i \neq 0$

$$\alpha_i = \frac{\mathbf{g}_i^T \mathbf{y}_i}{\mathbf{p}_i^T Q \mathbf{p}_i} \bullet$$

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

$$\mathbf{g}_{i+1} = \mathbf{g}_i + \alpha_i Q \mathbf{p}_i \bullet$$

$$\text{حل } M\mathbf{y}_{i+1} = \mathbf{g}_{i+1} \bullet$$

$$\beta_i = \frac{\mathbf{g}_{i+1}^T \mathbf{y}_{i+1}}{\mathbf{r}_i^T \mathbf{y}_i} \bullet$$

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

▪ به دنبال یافتن کمینه

$$M = S^T S \bullet$$

پیش شرطی

روش مستقیم بروز کردن x

$$W = S^T S$$

$$p_0 = W g_0$$

$$x_{k+1} = x_k + \alpha_k p_k$$

$$p_{k+1} = W g_{k+1} + \beta_k p_k$$

$$\beta_k = \frac{g_{k+1}^T W g_{k+1}}{g_k^T W g_k}$$

منابع

[نازهدل]

[لوئین برگر]

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>