بهینه سازی
مبانی بهینه سازی نامقید
گرادیان مزدوج

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

گرادیان مزدوج

هستنز و استیفل ۱۹۵۱

لنکزوس ۱۹۵۲

بهبود گرادیان نزولی با جلوگیری از تکرار مراحل

استفاده از تاریخچهء قدم قبلی جهت سریع کردن همگرائی

مشابه شدیدترین نزول به جز در محاسبهٔ کردن قدمها

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

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

حرکت در جهات متعامد

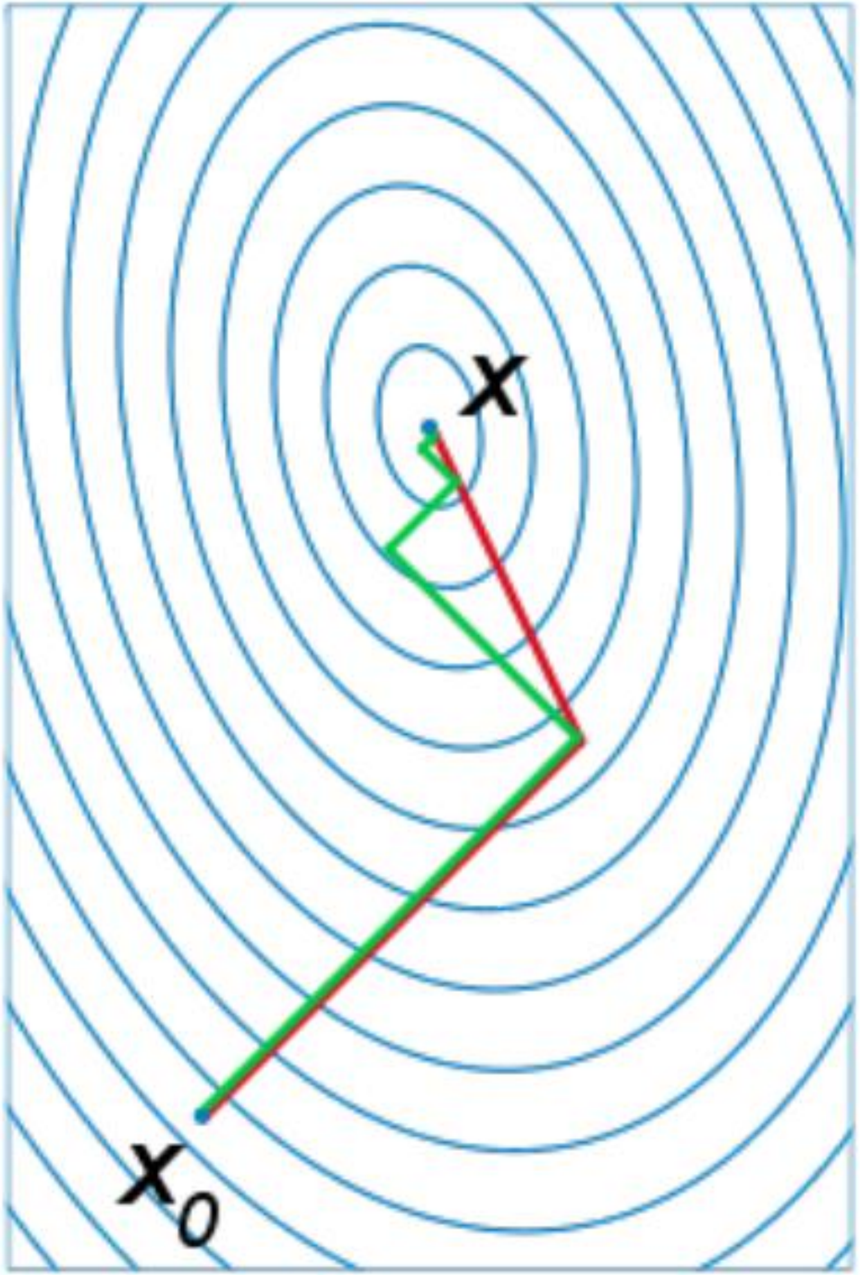
- جهت مزدوج با مسیر قبلی

متعامد یا مزدوجی دو بردار

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

جلوگیری از افزونگی

گرادیان مزدوج



هستنز و استیفل ۱۹۵۱

لنکزوس ۱۹۵۲

بهبود گرادیان نزولی با جلوگیری از تکرار مراحل

استفاده از تاریخچهء قدم قبلی جهت سریع کردن

مشابه شدیدترین نزول به جز در محاسبهٔ کردن ة

حرکت در جهات متعامد

- جهت مزدوج با مسیر قبلی

متعامد یا مزدوجی دو بردار

جلوگیری از افزونگی

ضرب داخلی

$$\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n \Longrightarrow \boldsymbol{a}.\boldsymbol{b} = \boldsymbol{a}^T\boldsymbol{b} = \sum a_i b_i$$

$$f, g \Longrightarrow \mathrm{f.g} = \int_T f(x)\bar{g}(\boldsymbol{x})dx$$

$\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, Q$ ماتریس مثبت معین $\Longrightarrow$

$$\Longrightarrow \boldsymbol{x}.\boldsymbol{y}_Q = \boldsymbol{x}^T Q\boldsymbol{y}$$

مسیرهای مزدوج

تعریف

Q ماتریس مثبت معین است. بردارهای $\{\boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_k\}$ با شرط $\boldsymbol{p}_i \in \mathbb{R}^n, \boldsymbol{p}_i \neq \boldsymbol{0}$ متعامد-Q (مزدوج-Q) نسبت به Q هستند اگر

$$\boldsymbol{p}_i^T Q \boldsymbol{p}_j = 0, \forall i \neq j$$

Q=0

Q=I

لم

Q مثبت معین باشد. اگر $\{\boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_k\}$ متعامد-Q باشند، آن‌گاه همگی مستقل خطی از یکدیگرند.

اثبات

اهمیت مزدوج-Q

هدف

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

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

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

فرض $\boldsymbol{x}^*$ پاسخ مسئله

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

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

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

اهمیت مزدوج-Q

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

پس

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

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

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

متعامدبودن معیار کافی نیست

- نیاز به متعامد-Q

اهمیت مزدوج-Q

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

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

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

صرفا احتیاج به ضرب داخلی

امکان یافتن نقطه کمینه‌سازی با استفاده از روش تکرار مراحل و عددی در **n** مرحله

- در مرحله **i**-ام مقدار $\alpha_i \boldsymbol{p}_i$ افزوده می‌شود
- تعمیم بیشتر با آغاز از نقطه دلبخواه $\boldsymbol{x}_0$

کمینه‌سازی توابع درجه دو

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

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

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

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

جداپذیر بر الفا

کمینه‌سازی توابع درجه دو

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

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

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

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

n دنباله از کمینه‌سازی‌های جستجو خط یک-بعدی

روش جهت مزدوج

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

n دنباله از کمینه‌سازی‌های جستجو خط یک-بعدی

بسط خاصیت زیرفضا- ادامه

$$\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{p}_i$$

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

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

پس از k مرحله از روش مزدوج

$$\boldsymbol{x}_k = \underset{\boldsymbol{x} \in G_k}{\operatorname{argmin}} f(\boldsymbol{x})$$

بسط خاصیت زیرفضا- ادامه

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

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

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

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

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

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

بسط خاصیت زیرفضا- ادامه

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

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

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

کمینه‌سازی توابع درجه دو

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

آغاز:

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

مرحلهٔ k-اُم

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

نیاز به جستجو خط کامل

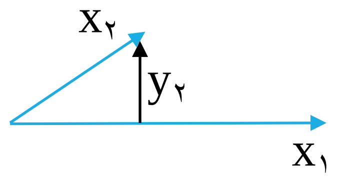
$$\boldsymbol{p}_{k+1} = -\boldsymbol{g}_{k+1} + \sum_{j=0}^{k}\left(\frac{{\boldsymbol{g}_{k+1}}^T Q\boldsymbol{p}_j}{{\boldsymbol{p}_j}^T Q\boldsymbol{p}_j}\boldsymbol{p}_j\right)$$

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

روش گرام اشمیت

$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3,\dots,x_j\}$ پایه‌ای برای X

- به دنبال پایه یکه متعامد

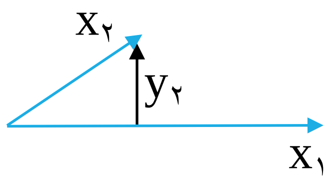


- $X_1 = span(\boldsymbol{x}_1) \Longrightarrow \boldsymbol{y}_1 = \frac{\boldsymbol{x}_1}{\|\boldsymbol{x}_1\|}$
- $X_2 = span(\boldsymbol{x}_1, \boldsymbol{x}_2) = span(\boldsymbol{y}_1, \boldsymbol{x}_2)$
- $\Longrightarrow \boldsymbol{y}_2 = \frac{\boldsymbol{x}_2 - (\boldsymbol{x}_2.\boldsymbol{y}_1)\boldsymbol{y}_1}{\|\boldsymbol{x}_2 - (\boldsymbol{x}_2.\boldsymbol{y}_1)\boldsymbol{y}_1\|} \Longrightarrow X_2 = span(\boldsymbol{y}_1, \boldsymbol{y}_2)$
- $X_3 = span(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = span(\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{x}_3)$
- $\Longrightarrow \boldsymbol{y}_3 = \frac{\boldsymbol{x}_3 - (\boldsymbol{x}_3.\boldsymbol{y}_1)\boldsymbol{y}_1 - (\boldsymbol{x}_3.\boldsymbol{y}_2)\boldsymbol{y}_2}{\|\boldsymbol{x}_3 - (\boldsymbol{x}_2.\boldsymbol{y}_1)\boldsymbol{y}_1 - (\boldsymbol{x}_3.\boldsymbol{y}_2)\boldsymbol{y}_2\|} \Longrightarrow X_3 = span(\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)$
- $\vdots$
- $X_j = span(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \dots, \boldsymbol{x}_j) = span(\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_{j-1}, \boldsymbol{x}_j)$
- $\Longrightarrow \boldsymbol{y}_j = \frac{\boldsymbol{x}_j - \sum_{i=1}^{j-1}(\boldsymbol{x}_j.\boldsymbol{y}_i)\boldsymbol{y}_i}{\left\|\boldsymbol{x}_j - \sum_{i=1}^{j-1}(\boldsymbol{x}_j.\boldsymbol{y}_i)\boldsymbol{y}_i\right\|} \Longrightarrow X_j = span(\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_j)$

متعامدسازی گرام-اشمیت- ادامه

بردارهای مستقل خطی $\{\boldsymbol{x}_۱,\boldsymbol{x}_۲, \dots ,\boldsymbol{x}_n\}$

ایجاد محورهای(پایه‌های) متعامد



$$\boldsymbol{y}_۱ = \boldsymbol{x}_۱$$

$$\boldsymbol{y}_۲ = \boldsymbol{x}_۲ - (\boldsymbol{x}_۲ . \frac{\boldsymbol{y}_۱}{\|\boldsymbol{y}_۱\|}) \frac{\boldsymbol{y}_۱}{\|\boldsymbol{y}_۱\|} = \boldsymbol{x}_۲ - \frac{\boldsymbol{x}_۲ . \boldsymbol{y}_۱}{\|\boldsymbol{y}_۱\|^۲} \boldsymbol{y}_۱$$

$$\vdots$$

$$\boldsymbol{y}_j = \boldsymbol{x}_j - \sum_{i=۱}^{j-۱} \frac{\boldsymbol{x}_j . \boldsymbol{y}_i}{\|\boldsymbol{y}_i\|^۲} \boldsymbol{y}_i$$

کمینه‌سازی توابع درجه دو

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

ساده‌سازی ۱

$$\boldsymbol{p}_j = \frac{1}{\alpha_j}\left(\boldsymbol{x}_{j+1} - \boldsymbol{x}_j\right) \Longrightarrow Q\boldsymbol{p}_j = \frac{1}{\alpha_j} Q\left(\boldsymbol{x}_{j+1} - \boldsymbol{x}_j\right) = \frac{1}{\alpha_j}(\boldsymbol{g}_{j+1} - \boldsymbol{g}_j)$$

$\boldsymbol{g}_{k+1}$ عمود بر $\boldsymbol{g}_0$، $\boldsymbol{g}_۱$،...، $\boldsymbol{g}_k$ ← پس آخرین بخش در جمع مقادیر بالای صفحه کافی است

کمینه‌سازی توابع درجه دو

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

ساده‌سازی ۱

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

ب) با قضیه بسط زیرفضا

$\boldsymbol{g}_{k+1}$ عمود بر $\boldsymbol{g}_0$، $\boldsymbol{g}_1$،....، $\boldsymbol{g}_k$ ← پس آخرین بخش کافی است

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

کمینه‌سازی توابع درجه دو

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

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

ساده‌سازی ۲

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

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

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

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

کمینه‌سازی توابع درجه دو

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

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

ساده‌سازی۲

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

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

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

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

پس

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

کمینه‌سازی توابع درجه دو

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

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

ساده‌سازی ۲

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

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

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

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

پس

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

کمینه‌سازی توابع درجه دو

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

پس

روش پولاک-ریبریه $\beta_k = \dfrac{{\boldsymbol{g}_{k+1}}^T(\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}{\|\boldsymbol{g}_k\|_2^2}$

روش فلچر-ریوز $\beta_k = \dfrac{\|\boldsymbol{g}_{k+1}\|_2^2}{\|\boldsymbol{g}_k\|_2^2}$

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

الگوریتم گرادیان مزدوج

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

- به دنبال یافتن کمینه

الگوریتم گرادیان مزدوج

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

برای n بار

- $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i$
- α_i کمینه‌ساز $g(\alpha) = f(\boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i)$
- $\boldsymbol{g}_{i+1} = \nabla f(\boldsymbol{x}_{i+1})$
- $\boldsymbol{p}_{i+1} = -\boldsymbol{g}_{i+1} + \beta_i \boldsymbol{p}_i$
- $\beta_i = \frac{{\boldsymbol{g}_{i+1}}^T(\boldsymbol{g}_{i+1} - \boldsymbol{g}_i)}{\|\boldsymbol{g}_i\|_2^2}$ یا $\beta_i = \frac{\|\boldsymbol{g}_{i+1}\|_2^2}{\|\boldsymbol{g}_i\|_2^2}$

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

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

TABLE 1.1
Various choices for the CG update parameter

روش مسیرهای مزدوج

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

آغاز:

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

مرحلهٔ **i**-اُم

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

نیاز به جستجو خط کامل

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

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

الگوریتم گرادیان مزدوج

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

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

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

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

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

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

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

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

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

$$\Rightarrow \alpha_i = -\frac{\boldsymbol{p}_i^T\boldsymbol{g}_i}{\boldsymbol{p}_i^T Q\boldsymbol{p}_i}$$

روش گرادیان مزدوج

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

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

آغاز:

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

مرحلهٔ **i**-اُم

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

نیاز به جستجو خط کامل

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

روش گرادیان مزدوج

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

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

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

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

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

الگوریتم گرادیان مزدوج

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

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

- به دنبال یافتن کمینه

الگوریتم گرادیان مزدوج۲

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

برای n بار

- $\alpha_i = -\frac{\boldsymbol{g}_i{}^T\boldsymbol{p}_i}{\boldsymbol{p}_i{}^T Q\boldsymbol{p}_i}$
- $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i\boldsymbol{p}_i$
- $\boldsymbol{g}_{i+1} = Q\boldsymbol{x}_{i+1} - \boldsymbol{b}$
- $\beta_i = \frac{\boldsymbol{g}_{i+1}{}^T Q\boldsymbol{p}_i}{\boldsymbol{p}_i{}^T Q\boldsymbol{p}_i}$
- $p_{i+1} = -\boldsymbol{g}_{i+1} + \beta_i\boldsymbol{p}_i$

مثال

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

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

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

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

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

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

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

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

$$p_1=-\boldsymbol{g}_1+\beta_0p_0=-\begin{bmatrix}0\\-\frac{1}{2}\end{bmatrix}+\frac{1}{4}\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}\frac{1}{4}\\\frac{1}{2}\end{bmatrix}$$

$$\alpha_1=-\frac{\boldsymbol{g}_1{}^T\boldsymbol{p}_1}{\boldsymbol{p}_1{}^TQ\boldsymbol{p}_1}=-\frac{\begin{bmatrix}0 & -\frac{1}{2}\end{bmatrix}\begin{bmatrix}\frac{1}{4}\\\frac{1}{2}\end{bmatrix}}{\begin{bmatrix}\frac{1}{4} & \frac{1}{2}\end{bmatrix}\begin{bmatrix}2 & -1\\ -1 & 2\end{bmatrix}\begin{bmatrix}\frac{1}{4}\\\frac{1}{2}\end{bmatrix}}=\frac{2}{3}$$

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

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

الگوریتم گرادیان مزدوج روش کاراتر

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

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

- به دنبال یافتن کمینه

الگوریتم گرادیان مزدوج۳

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

تازمان $\boldsymbol{g}_i \neq 0$

- $\alpha_i = \dfrac{{\boldsymbol{g}_i}^T\boldsymbol{g}_i}{{\boldsymbol{p}_i}^T Q\boldsymbol{p}_i}$
- $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i\boldsymbol{p}_i$
- $\boldsymbol{g}_{i+1} = \boldsymbol{g}_i + \alpha_i Q\boldsymbol{p}_i$
- $\beta_i = \dfrac{{\boldsymbol{g}_{i+1}}^T\boldsymbol{g}_{i+1}}{{\boldsymbol{g}_i}^T\boldsymbol{g}_i}$
- $\boldsymbol{p}_{i+1} = -\boldsymbol{g}_{i+1} + \beta_i\boldsymbol{p}_i$

مثال

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

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

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

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

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

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

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

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

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

$$p_1 = -\boldsymbol{g}_1 + \beta_0 p_0 = \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix} + \frac{1}{4}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \end{bmatrix}$$

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

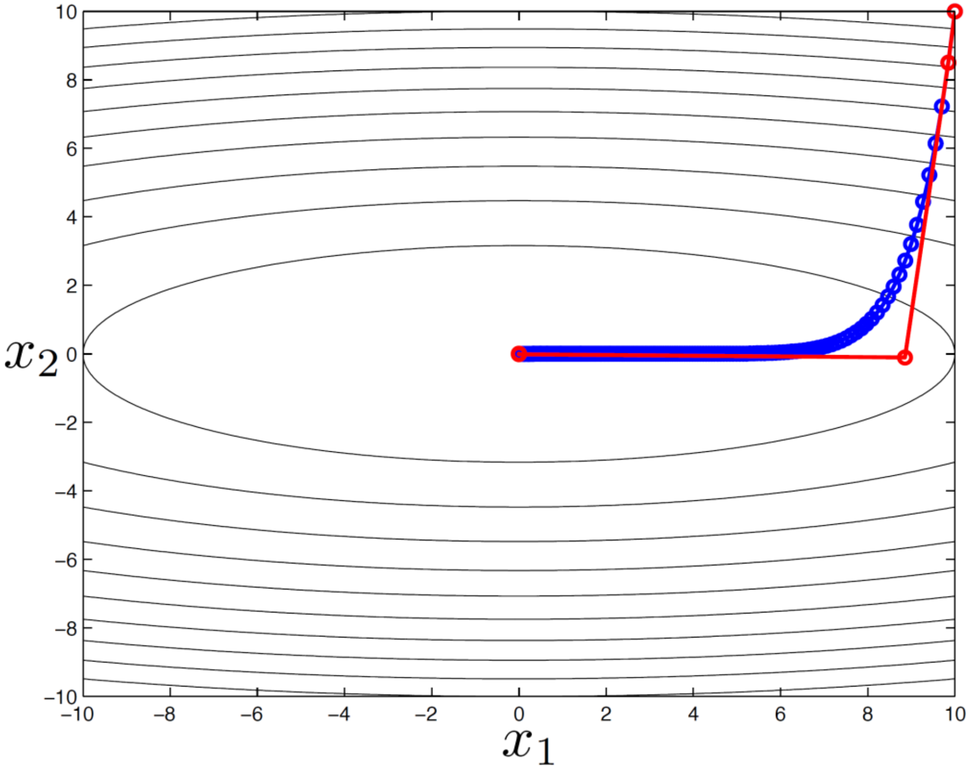
$$\boldsymbol{x}_2 = \boldsymbol{x}_1 + \alpha_1 \boldsymbol{p}_1 = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} + \frac{2}{3}\begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$$

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

مثال- تندترین نزول در برابر گرادیان مزدوج

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

$$\alpha_i = 0.015$$



گرادیان مزدوج

«تصحیح» جهت شدیدترین نزول

$$\boldsymbol{p}_i = -\nabla f(x_i) + \beta_i \boldsymbol{p}_{i-1}$$

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

- اجازهٔ استفاده از جهت قبلی

مزایا

- کاراتر از گردایان کاهشی
- پیاده‌سازی ساده به همان میزان
- صرفا اطلاع مرتبهٔ اول
- حافظهٔ کم
- قابل استفاده جهت توابع غیرخطی

معایب

- سرعت همگرایی متوسط
- حساس به مقیاس‌گذاری

سرعت همگرائی گرادیان مزدوج

$$0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$$

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

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

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

سرعت همگرائی گرادیان مزدوج

$$\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^*\| \le \frac{\sqrt{\lambda_n}-\sqrt{\lambda_1}}{\sqrt{\lambda_n}+\sqrt{\lambda_1}}\|\boldsymbol{x}_k-\boldsymbol{x}^*\|$$

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

پیش‌شرطی

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

تغییر متغیرها

- فرض S معکوس‌پذیر و $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x}$

جایگذاری در تابع هدف

$$\phi(y) = \frac{1}{2}\boldsymbol{y}^T(S^{-T}QS^{-1})\boldsymbol{y} - (S^{-T}\boldsymbol{b})^T\boldsymbol{y}$$

$$(S^{-T}QS^{-1})\boldsymbol{y} = (S^{-T}\boldsymbol{b})^T$$

سرعت همگرائی به جای Q وابسته به مقداروی‌ژه‌های ماتریس $S^{-T}QS^{-1}$

- پس به دنبال S-ی که مقدار ویژه‌های مناسبتر جهت همگرائی
  - تبدیل به ماتریسی با عدد شرط کمتر

الگوریتم گرادیان مزدوج پیش‌شرطی

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

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

- به دنبال یافتن کمینه
- $M = S^T S$

الگوریتم گرادیان مزدوج پیش‌شرطی

M پیش‌شرط‌ساز و

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

حل $M\boldsymbol{y}_0 = \boldsymbol{g}_0$

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

تازمان $\boldsymbol{g}_i \neq 0$

- $\alpha_i = \dfrac{\boldsymbol{g}_i{}^T \boldsymbol{y}_i}{\boldsymbol{p}_i{}^T Q \boldsymbol{p}_i}$
- $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \alpha_i \boldsymbol{p}_i$
- $\boldsymbol{g}_{i+1} = \boldsymbol{g}_i + \alpha_i Q \boldsymbol{p}_i$
- حل $M\boldsymbol{y}_{i+1} = \boldsymbol{g}_{i+1}$
- $\beta_i = \dfrac{\boldsymbol{g}_{i+1}{}^T \boldsymbol{y}_{i+1}}{\boldsymbol{g}_i{}^T \boldsymbol{y}_i}$
- $p_{i+1} = -\boldsymbol{y}_{i+1} + \beta_i p_i$

پیش‌شرطی

روش مستقیم بروزکردن x

$$W = S^T S$$

$$\boldsymbol{p}_0 = W\boldsymbol{g}_0$$

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

$$\boldsymbol{p}_{k+1} = W\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{p}_k$$

$$\beta_k = \frac{\boldsymbol{g}_{k+1}^T W \boldsymbol{g}_{k+1}}{\boldsymbol{g}_k^T W \boldsymbol{g}_k}$$

منابع

[نازه‌دل]

[لوئن‌برگر]

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