بهینه سازی
روش‌های منطقه اعتماد

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

Actual reduction

Cauchy point

Conjugate gradient

Curved trajectory

method

Dogleg method

Restricted step

Predicted reduction

Trust region

Trust region radius

Trust region step

Two-dimensional subspace minimization

بهینه‌سازی نامقید- جستجو خط و منطقه اعتماد

استراتژی جستجو خط

- انتخاب جهت $\boldsymbol{p}_i$
- جستجو در راستای جهت یافت شده از نقطهٔ $\boldsymbol{x}_i$ فعلی به نقطه‌ای دیگر با مقدار f کمتر
- با حل مسئله کمینه‌سازی یک‌بعدی زیر جهت یافتن α

$$\underset{\alpha>0}{\text{کم}} f(\boldsymbol{x}_i + \alpha \boldsymbol{p}_i)$$

استراتژی منطقه اعتماد

- تعریف شعاعی $\Delta > 0$ اطراف $\boldsymbol{x}_i$ به عنوان منطقهٔ اعتماد
- جمع‌آوری اطلاعات جهت ایجاد مدلی (تخمینی) از تابع f
  - تابع تخمین با نام m_i
- دارای رفتار مشابه در نزدیکی نقطه $\boldsymbol{x}_i$

$$\underset{\boldsymbol{p}}{\text{کم}} m_i(\boldsymbol{p}_i)$$

$\boldsymbol{p}$ در داخل منطقه اعتماد

- $\|\boldsymbol{p}\|_2 \le \Delta$

روش نضج

تقریب تابع f با مدل ساده‌تر m

$$\hat{\boldsymbol{x}}^* = \underset{\boldsymbol{x}}{argmin}\, m(\boldsymbol{x})$$

مشکل

- $\hat{\boldsymbol{x}}^*$ در جایی که m تقریب نامناسب و ضعیفی از f باشد

حل

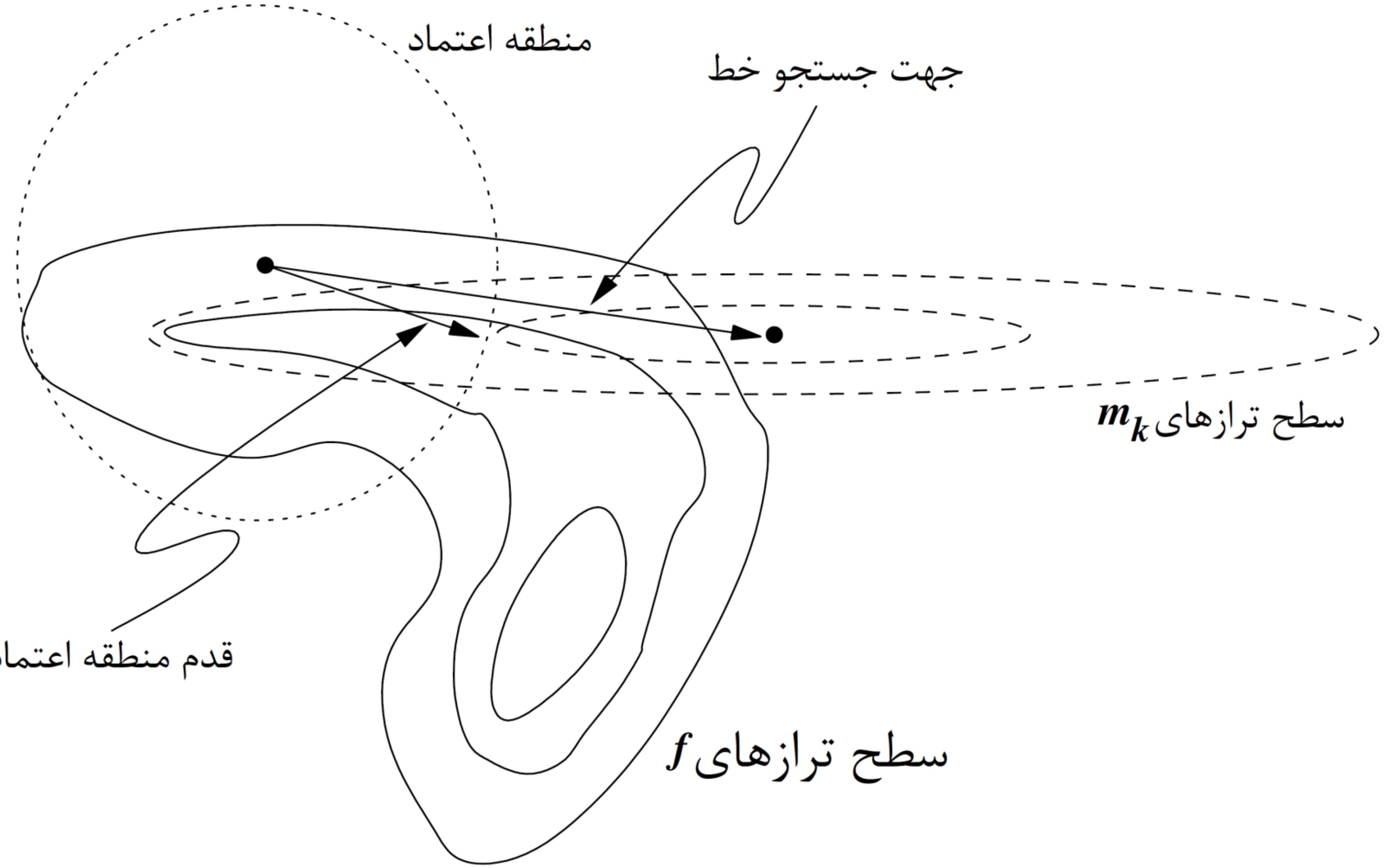
- محدود کردن جستجو به ناحیه‌ای که m تقریب مناسبی از f باشد
  - منطقه اعتماد Ω

$$\hat{\boldsymbol{x}}^* = \underset{\boldsymbol{x}\in\Omega}{argmin}\, m(\boldsymbol{x})$$

روش منطقه اعتماد

قدم محدودشده

نضج روش


منطقه اعتماد
جهت جستجو خط
سطح ترازهای m_k
قدم منطقه اعتماد
سطح ترازهای f

استراتژی منطقه اعتماد

تابع تخمین m_i معمولا برابر با

$$m_i(\boldsymbol{p}_i) = f_i + \boldsymbol{p}_i{}^T \nabla f_i + \frac{1}{2}\boldsymbol{p}_i{}^T B_i \boldsymbol{p}_i$$

m_i بسط سری تیلور f

B_i یا تابع هسی $\nabla^2 f_i$ یا تخمینی از آن

مقایسه با

$$f(\boldsymbol{x}_i + \boldsymbol{p}) \approx f(\boldsymbol{x}_i) + \nabla f(\boldsymbol{x}_i)^T \boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^T \nabla^2 f(\boldsymbol{x}_i + t\boldsymbol{p})\boldsymbol{p}$$

- دارای اختلاف $O(\|\boldsymbol{p}\|^2)$
- اگر ماتریس هسی دارای اختلاف $O(\|\boldsymbol{p}\|^3)$
- برای مقادیر کوچک $\mathbf{p}$ مدل نسبتا درست

استراتژی منطقه اعتماد

$$\min_{\boldsymbol{p}_i \in \mathbb{R}^n} m_i(\boldsymbol{p}_i) = f_i + \boldsymbol{p}_i{}^T \nabla f_i + \frac{1}{2}\boldsymbol{p}_i{}^T B_i \boldsymbol{p}_i$$

به‌طوری‌که

$$\|\boldsymbol{p}_i\| \le \Delta_i$$

Δ_i شعاع منطقه اعتماد

نیاز به چندین بار حل معادله برای رسیدن به کمینه

اگر B_i مثبت معین و $\left\|B_i{}^{-1}\boldsymbol{g}_i\right\| \le \Delta_i$

- راحتی یافتن پاسخ معادله
- $\boldsymbol{p}_i^B$
- قدم کامل
- $\boldsymbol{p}_i^B = -B_i{}^{-1}\boldsymbol{g}_i$

استراتژی منطقه اعتماد

از اجزای اصلی

- انتخاب شعاع مناسب منطقه در هر مرحله

انتخاب بر اساس تناسب بین مدل و تابع هدف در قدم‌های قبلی

با داشتن $\boldsymbol{p}_i$

$$\rho_i = \frac{f(\boldsymbol{x}_i) - f(\boldsymbol{x}_i + \boldsymbol{p}_i)}{m_i(\boldsymbol{0}) - m_i(\boldsymbol{p}_i)}$$

کاهش حقیقی $f(\boldsymbol{x}_i) - f(\boldsymbol{x}_i + \boldsymbol{p}_i)$

کاهش پیش‌بینی‌شده $m_i(\boldsymbol{0}) - m_i(\boldsymbol{p}_i)$

استراتژی منطقه اعتماد

$$\rho_i = \frac{f(\boldsymbol{x}_i) - f(\boldsymbol{x}_i + \boldsymbol{p}_i)}{m_i(0) - m_i(\boldsymbol{p}_i)}$$

کاهش حقیقی $f(\boldsymbol{x}_i) - f(\boldsymbol{x}_i + \boldsymbol{p}_i)$

کاهش پیش‌بینی‌شده $m_i(0) - m_i(\boldsymbol{p}_i)$

نسبت نزدیک ۱ (مقدار بزرگ)

- توافق مناسب بین واقع و تقریب
- به دیگر سخن، کمتر شدن مقدار واقعی بهتر از مقدار پیش‌بینی‌شده
- پیش‌بینی کم‌تر از مقدار واقع
- همه‌چیز مهیا برای افزایش منطقه اطمینان در مرحلهٔ بعد

نسبت مثبت ولی کوچکتر از ۱

- عدم تغییر منطقه اعتماد

نسبت نزدیک صفر یا منفی

- کوچک کردن شعاع در مرحلهٔ بعد

الگوریتم منطقه اعتماد

مشخص کردن مقادیر اولیه $\widehat{\Delta} > 0$ و $\Delta_0 \in (0,\widehat{\Delta})$ و $\eta \in [0,0.25)$

برای $i = 1:n$

- محاسبه $\boldsymbol{p}_i$ از $\min_{\boldsymbol{p}_i \in \mathbb{R}^n} m_i(\boldsymbol{p}_i) = f_i + \boldsymbol{p}_i^T \nabla f_i + \frac{1}{2}\boldsymbol{p}_i^T B_i \boldsymbol{p}_i$
- $\rho_i = \frac{f(\boldsymbol{x}_i) - f(\boldsymbol{x}_i + \boldsymbol{p}_i)}{m_i(0) - m_i(\boldsymbol{p}_i)}$
- اگر $\rho_i < ۰.۲۵$
  - $\Delta_{i+1} = \frac{۱}{۴}\Delta_i$
- وگرنه
  - اگر $\rho_i > ۰.۷۵$ و $\|\boldsymbol{p}_i\| = \Delta_i$
    - $\Delta_{i+1} = \min(2\Delta_i, \widehat{\Delta})$
  - وگرنه
    - $\Delta_{i+1} = \Delta_i$
- اگر $\rho_i > \eta$
  - $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \boldsymbol{p}_i$
- وگرنه
  - $\boldsymbol{x}_{i+1} = \boldsymbol{x}_i$

الگوریتم منطقه اعتماد

$\hat{\Delta}$ محدودهٔ بیشینه روی طول قدم‌ها

افزایش شعاع صرفا در صورت رسیدن $\|\boldsymbol{p}_i\|$ به مرزهای منطقه اعتماد

عدم تغییر شعاع در صورت درون مرز باقی‌ماندن

شروط یافتن $\boldsymbol{p}_i^*$ در $\min_{\boldsymbol{p}_i \in \mathbb{R}^n} m_i(\boldsymbol{p}_i) = f_i + {\boldsymbol{p}_i}^T \nabla f_i + \frac{1}{2} {\boldsymbol{p}_i}^T B_i \boldsymbol{p}_i$

- $\boldsymbol{p}_i^*$ شدنی
- $(B + \lambda I)$ مثبت معین

$$(B + \lambda I) p^* = -g,$$

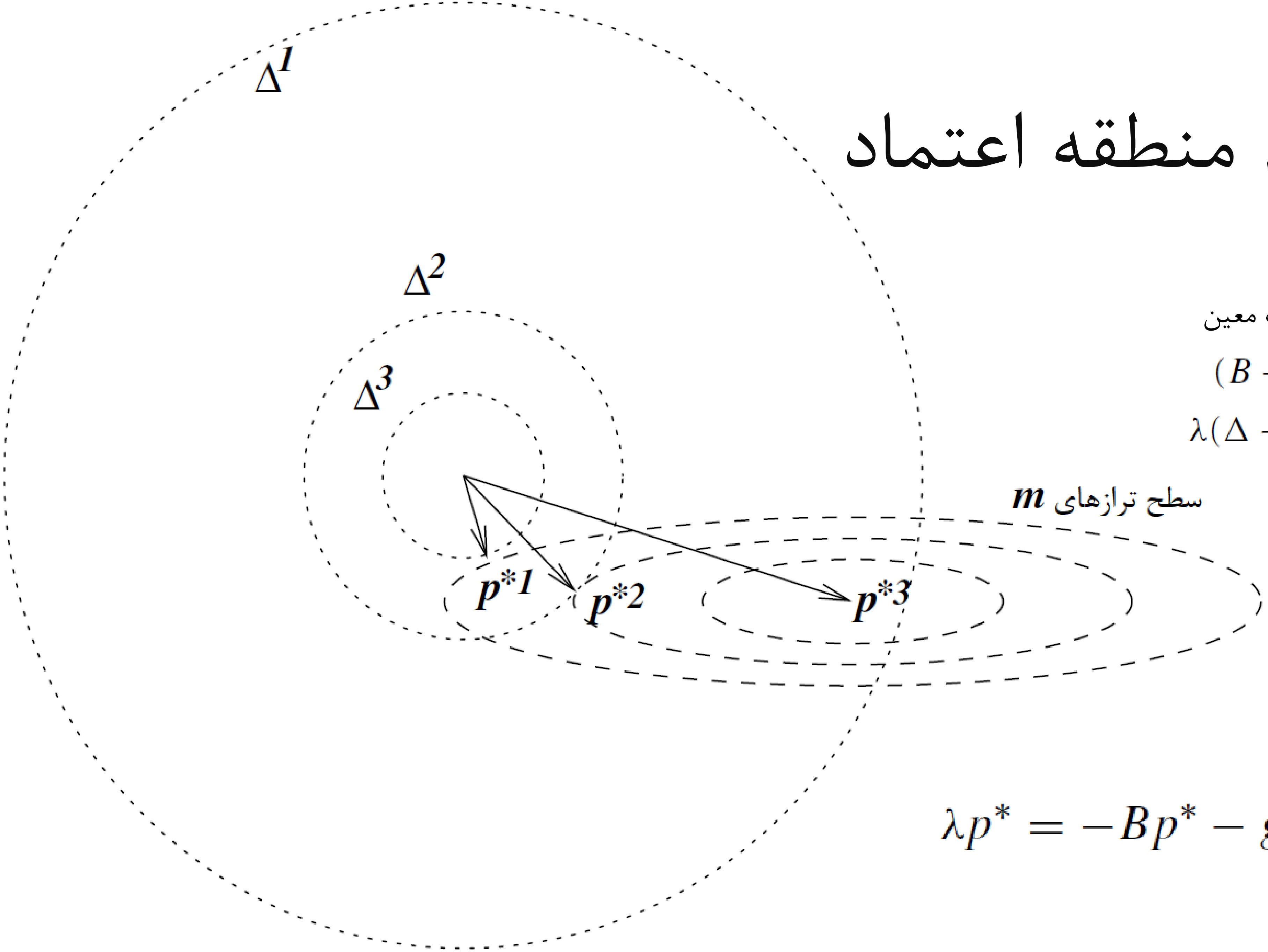
$$\lambda(\Delta - \|p^*\|) = 0,$$

الگوریتم منطقه اعتماد

- $(B + \lambda I)$ مثبت معین

$$(B+\lambda I)p^* = -g,$$

$$\lambda(\Delta - ||p^*||) = 0,$$



$$\lambda p^* = -Bp^* - g = -\nabla m(p^*)$$

نقطه کوشی

امکان همگرایی بدون نیاز به انتخاب بهترین طول قدم

- ؟

به طریق مشابه در روش منطقه اعتماد

- صرفا به دنبال تقریبی از $\boldsymbol{p}_i$ در منطقه اعتماد
  - انجام کاهش کافی
  - تدوین در قالب نقطه کوشی
  - نمایش با $\boldsymbol{p}_i^c$

الگوریتم محاسبهٔ نقطه کوشی

محاسبه $\boldsymbol{p}_i^s$ با حل معادله خطی

$$\boldsymbol{p}_i^s = \underset{\boldsymbol{p}_i \in \mathbb{R}^n}{argmin}\, f_i + {\boldsymbol{p}_i}^T \nabla f_i$$

به‌طوری‌که

$$\|\boldsymbol{p}_i\| \leq \Delta_i$$

محاسبه $\tau_i > 0$ کمینه‌کننده $m_i(\tau_i \boldsymbol{p}_i^s)$

- با رعایت شرط ماندن در منطقه اعتماد:

$$\tau_i = \underset{\tau > 0}{argmin}\, m_i(\tau \boldsymbol{p}_i^s)$$

به‌طوری‌که

$$\|\tau \boldsymbol{p}_i^s\| \leq \Delta_i$$

$$\boldsymbol{p}_i^c = \tau_i \boldsymbol{p}_i^s$$

الگوریتم محاسبهٔ نقطه کوشی

محاسبه $\boldsymbol{p}_i^s$ با حل معادله خطی

$$\boldsymbol{p}_i^s = \underset{\boldsymbol{p}_i \in \mathbb{R}^n}{argmin}\, f_i + {\boldsymbol{p}_i}^T \nabla f_i$$

به‌طوری‌که

$$\|\boldsymbol{p}_i\| \leq \Delta_i$$

محاسبه $\boldsymbol{p}_i^s$ با تدوین مستقیم

$$\boldsymbol{p}_i^s = -\frac{\Delta_i}{\|\boldsymbol{g}_i\|}\boldsymbol{g}_i$$

الگوریتم محاسبهٔ نقطه کوشی

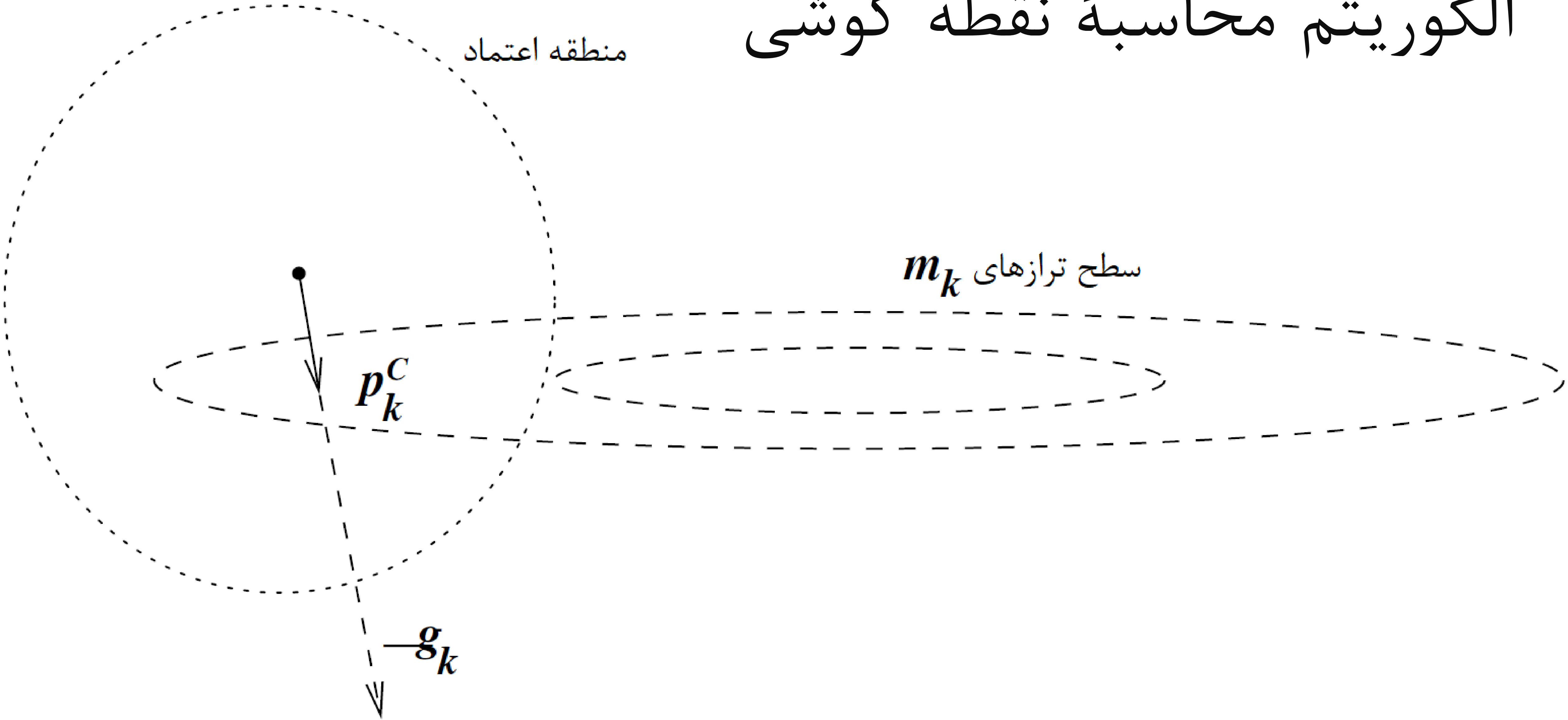
محاسبه $\boldsymbol{p}_i^s$ با تدوین مستقیم

$$\boldsymbol{p}_i^s = -\frac{\Delta_i}{\|\boldsymbol{g}_i\|}\boldsymbol{g}_i$$

محاسبه τ_i

$$\tau_i = \begin{cases} 1, & \boldsymbol{g}_i^T B_i \boldsymbol{g}_i \le 0 \\ \min\left(\dfrac{\|\boldsymbol{g}_i\|^3}{\Delta_i \boldsymbol{g}_i^T B_i \boldsymbol{g}_i}, 1\right) & \boldsymbol{g}_i^T B_i \boldsymbol{g}_i > 0 \end{cases}$$

الگوریتم محاسبهٔ نقطه کوشی



الگوریتم محاسبهٔ نقطه کوشی

محاسبات کم

بدون نیاز به محاسبات ماتریس

مناسب جهت تصمیم‌گیری درباره پذیرش حل تقریبی مسئله منطقه اعتماد

الگوریتم محاسبهٔ نقطه کوشی

مشابه تندترین نزول کند

چند روش بهبود

مثال

- استفاده از قدم کامل $\boldsymbol{p}_i^B = -{B_i}^{-1}\boldsymbol{g}_i$ هنگام
  - B_i مثبت معین
  - $\left\|\boldsymbol{p}_i^B\right\| \le \Delta_i$
  - در صورت هسی یا یا تقریب شبه نیوتنی
    - همگرائی زبرخطی

روش پاسگ

قابل استفاده هنگام مثبت معین بودن B

اگر $\|\boldsymbol{p}_i^B\| \leq \Delta_i$

- $p^*(\Delta_i) = \boldsymbol{p}_i^B$

اگر $\|\boldsymbol{p}_i^B\| > \Delta_i$

- حذف بخش مرتبه دو

با کوچک بودن Δ_i

$$p^*(\Delta) \approx -\Delta \frac{g}{\|g\|},$$

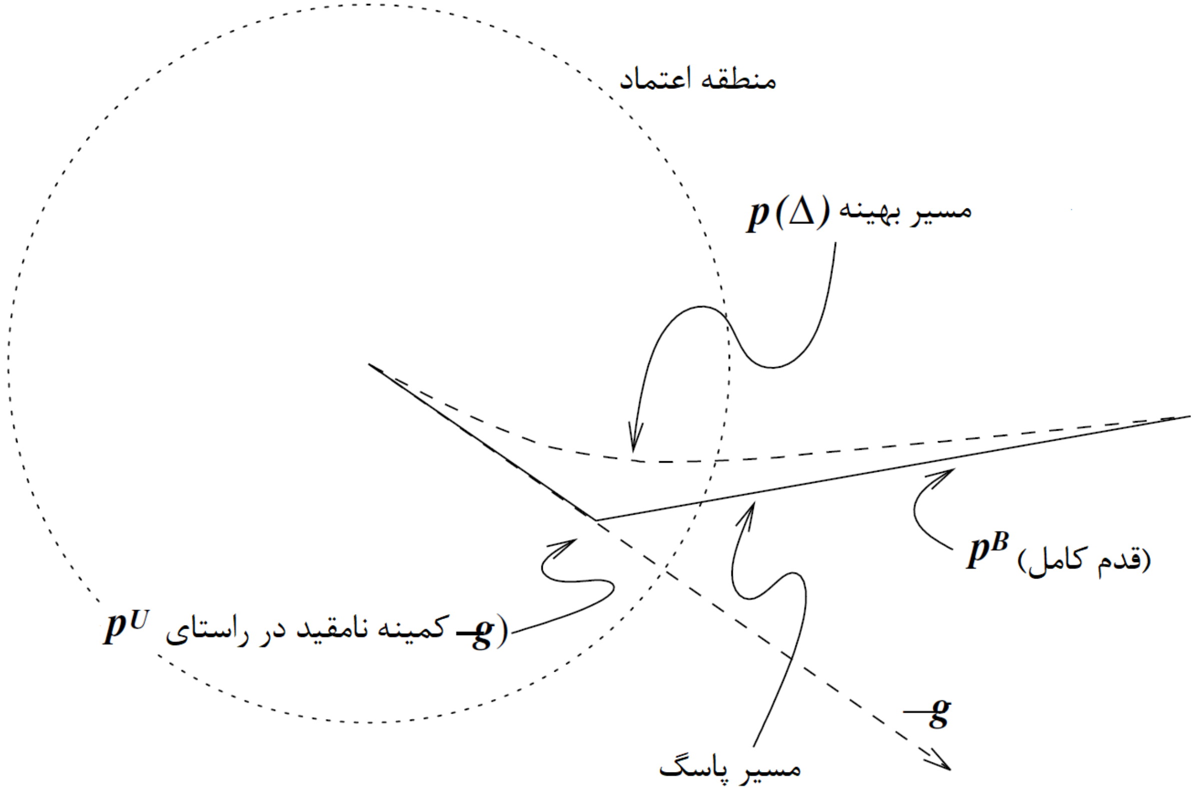
روش پاسگ

اگر $\Delta_i < \|\boldsymbol{p}_i^B\|$

با کوچک بودن Δ_i
$$p^*(\Delta) \approx -\Delta \frac{g}{\|g\|},$$

اگر Δ_i مقادیر وسط

روش پاسگ



اگر $\|p_i^B\| > \Delta_i$

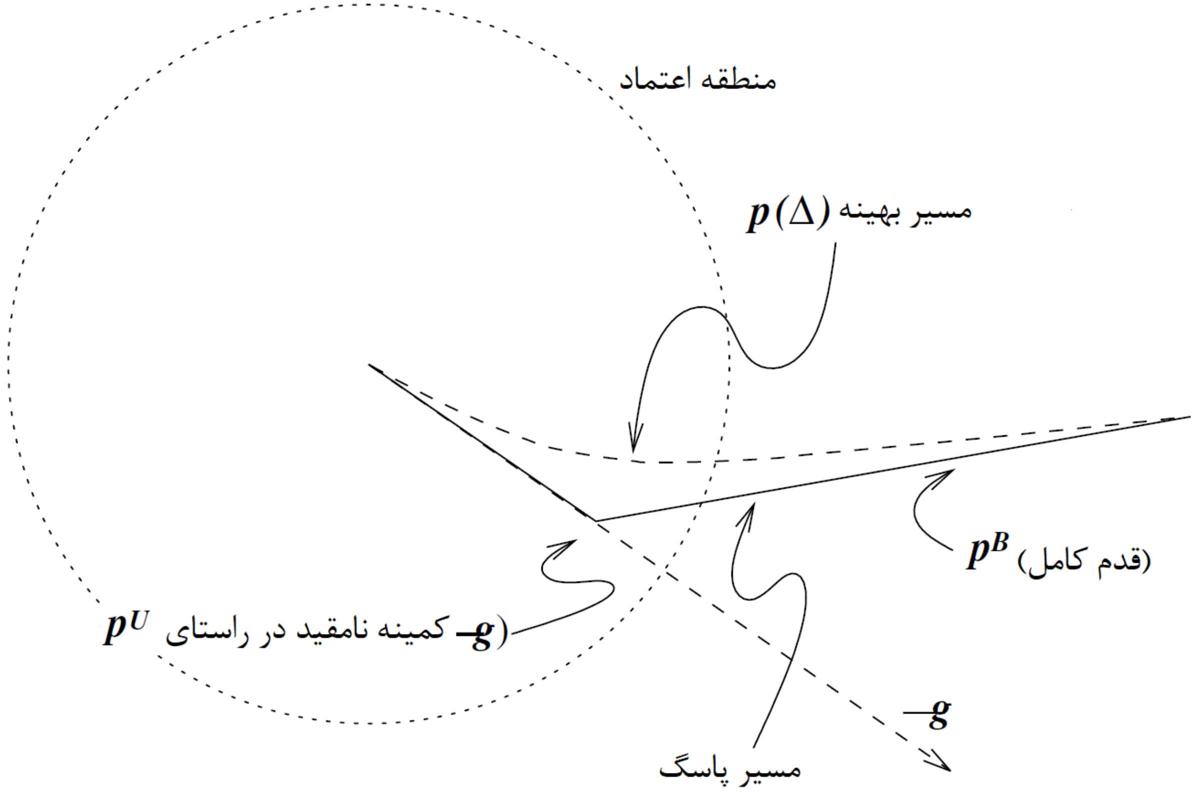
با کوچک بودن Δ_i

$$\Delta) \approx -\Delta \frac{g}{\|g\|},$$

اگر Δ_i مقادیر وسط

- جانشینی مسیر منحنی با دو خط مستقیم به دنبال هم
  - نخستین خط از مبدا به در راستای کمترین نزول

$$p^{\mathrm{U}} = -\frac{g^T g}{g^T B g} g,$$



روش پاسگ

اگر $\|\boldsymbol{p}_i^B\| > \Delta_i$

با کوچک بودن Δ_i

$$(\Delta) \approx -\Delta \frac{g}{\|g\|},$$

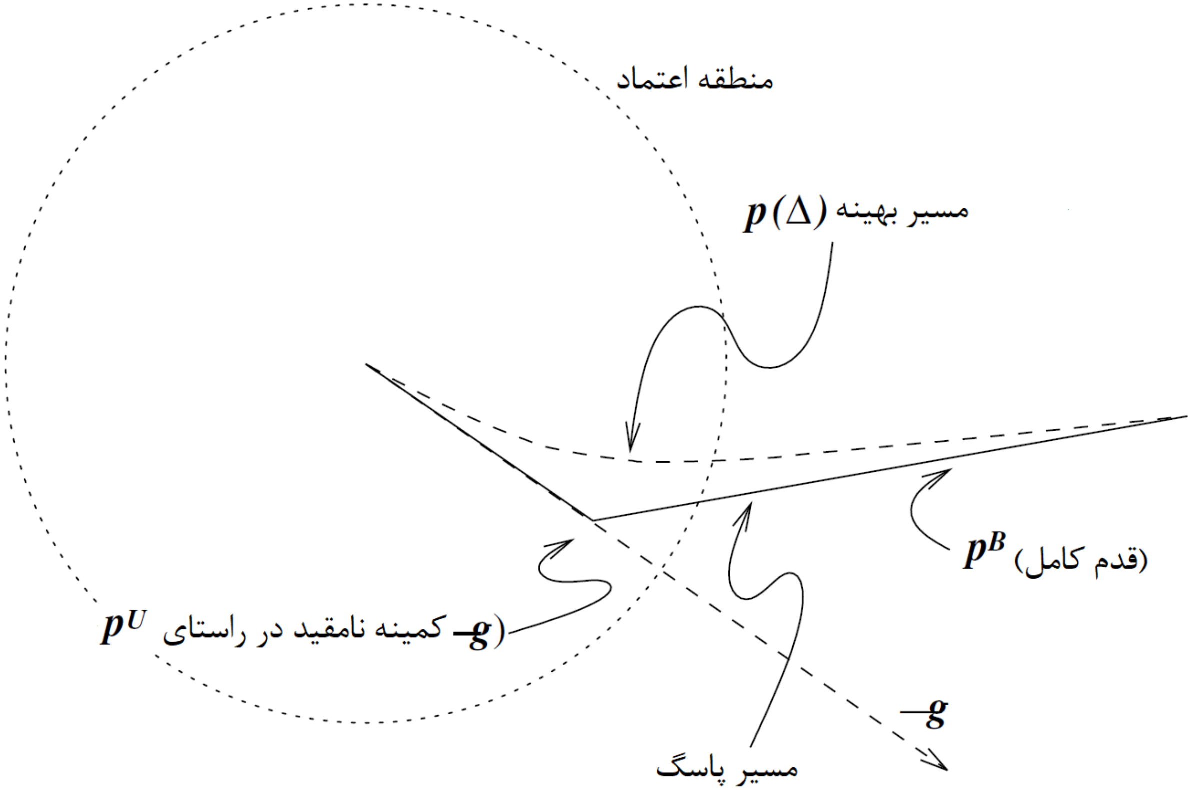
اگر Δ_i مقادیر وسط

- خط اول

$$p^{\mathrm{U}} = -\frac{g^T g}{g^T B g} g,$$

- خط دوم در راستای قدم کامل

روش پاسگ



اگر $\|p_i^B\| > \Delta_i$

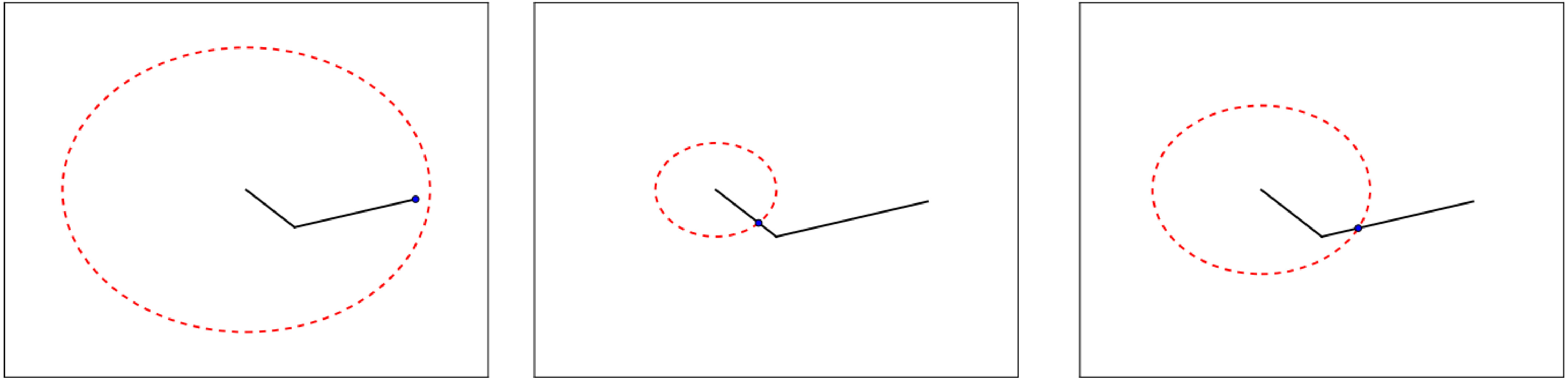
با کوچک بودن Δ_i

$$\Delta) \approx -\Delta \frac{g}{\|g\|},$$

اگر Δ_i مقادیر وسط

$$\tilde{p}(\tau) = \begin{cases} \tau p^{\mathrm{U}}, & 0 \le \tau \le 1, \\ p^{\mathrm{U}} + (\tau - 1)(p^{\mathrm{B}} - p^{\mathrm{U}}), & 1 \le \tau \le 2. \end{cases}$$

روش پاسگ



$$\tilde{p}(\tau) = \begin{cases} \tau p^{U}, & 0 \leq \tau \leq 1, \\ p^{U} + (\tau - 1)(p^{B} - p^{U}), & 1 \leq \tau \leq 2. \end{cases}$$

http://www.acme.byu.edu/wp-content/uploads/2014/09/Vol2Lab19TrustRegion.pdf

الگوریتم منطقه اعتماد

آستانه $\epsilon_i > 0$

مشخص کردن مقادیر اولیه $z_0 = 0$ و $\mathbf{g}_0 = \nabla f_i$ و $\mathbf{d}_0 = -\boldsymbol{g}_0 = -\nabla f_i$

اگر $\|\boldsymbol{g}_0\| < \epsilon_i$

- $\boldsymbol{p}_i = \mathbf{z}_0 = 0$ و خروج

برای $j = 0,1, \dots$

}

- اگر $\boldsymbol{d}^T B_i \boldsymbol{d} \leq 0$
  - یافتن به طوری که $\boldsymbol{p}_i = \boldsymbol{z}_j + \tau \mathbf{d}_j$ با رعایت دو شرط
    - کمینه‌ساز $\min_{\boldsymbol{p}_i \in \mathbb{R}^n} m_i(\boldsymbol{p}_i) = f_i + \boldsymbol{p}_i{}^T \nabla f_i + \frac{1}{2} \boldsymbol{p}_i{}^T B_i \boldsymbol{p}_i$
    - $\|\boldsymbol{p}_i\| == \Delta_i$
  - برگرداندن $\boldsymbol{p}_i$
- $\alpha_j = \boldsymbol{g}_j^T \boldsymbol{g}_j / \boldsymbol{d}_j^T B_i \boldsymbol{d}_j$
- $\boldsymbol{z}_{j+1} = \boldsymbol{z}_j + \alpha_j \boldsymbol{d}_j$
- اگر $\|\boldsymbol{z}_{j+1}\| \geq \Delta_i$
  - یافتن به طوری که $\boldsymbol{p}_i = \boldsymbol{z}_j + \tau \mathbf{d}_j$ با رعایت شرط $\|\boldsymbol{p}_i\| == \Delta_i$ و برگرداندن $\boldsymbol{p}_i$
- $\boldsymbol{g}_{j+1} = \boldsymbol{g}_j + \alpha_j B_i \boldsymbol{d}_j$
- اگر $\|\boldsymbol{g}_{j+1}\| < \epsilon_i$
  - برگرداندن $\boldsymbol{p}_i = \boldsymbol{z}_{j+1}$
- $\beta_{j+1} = \boldsymbol{g}_{j+1}^T \boldsymbol{g}_{j+1} / \boldsymbol{g}_j^T \boldsymbol{g}_j$
- $\boldsymbol{d}_{j+1} = -\boldsymbol{g}_{j+1} + \beta_{j+1} \boldsymbol{d}_j$

{

منابع

[نازەدل]

[فلچر]