Villkor på minimerare

Antag x_* lokal minimerare till f. Studera f runt x_*:

$$f(x_* + p) = f(x_*) + \nabla f(x_*)^T p + \frac{1}{2} p^T \nabla^2 f(\xi) p$$

Om x_* lokal minimerare så finns ingen tillåten nedförsriktning, dvs

$$\nabla f(x_*)^T p \geq 0$$

för alla tillåtna riktningar p.

För problem utan bivillkor medför detta att

$$\nabla f(x_*) = 0 \qquad (1)$$

En punkt som uppfyller (1) kallas *stationär punkt* till f.

Villkoret (1) kallas för ett *nödvändigt* villkor av första graden för en optimerare.

Villkor på minimerare, forts.

Studera f runt stationär punkt x_* som antas vara en lokal minimerare:

$$f(x) = f(x_* + p) = f(x_*) + \underbrace{\nabla f(x_*)^T p}_{=0} + \frac{1}{2} p^T \nabla^2 f(\xi) p.$$

För x nära x_* kommer $\nabla^2 f(\xi)$ vara nära $\nabla^2 f(x_*)$.

Om $\nabla^2 f(x_*)$ ej positivt semi-definit finns v sådan att

$$v^T \nabla^2 f(x_*) v < 0.$$

Då finns p nära v sådan att

$$p^T \nabla^2 f(\xi) p < 0$$

Villkor på minimerare, forts.

vilket medför att

$$\begin{aligned} f(x) &= f(x_* + p) = f(x_*) + \underbrace{\frac{1}{2} p^T \nabla^2 f(\xi) p}_{<0} \\ f(x) &< f(x_*) \end{aligned}$$

vilket är en motsägelse, ty x_* antogs vara en minimerare.

Alltså måste $\nabla^2 f(x_*)$ vara positivt semi-definit för att en stationär punkt x_* ska vara en lokal minimerare.

Detta kallas för ett *nödvändigt* minimeringsvillkor av andra graden.

Villkor på minimerare, forts.

Studera f runt x_* då $\nabla f(x_*) = 0$ och $\nabla^2 f(x_*)$ positivt definit.

Då är

$$f(x) = f(x_* + p) = f(x_*) + \underbrace{\nabla f(x_*)^T p}_{=0} + \frac{1}{2} p^T \nabla^2 f(\xi) p.$$

För x tillräckligt nära x_* kommer $\nabla^2 f(\xi)$ också att vara positivt definit, dvs

$$\begin{aligned} f(x) &= f(x_* + p) = f(x_*) + \underbrace{\frac{1}{2} p^T \nabla^2 f(\xi) p}_{>0 \ \forall p \neq 0} \\ f(x) &> f(x_*) \quad \forall p \end{aligned}$$

Vilket innebär att x_* är en strikt minimerare till f.

Detta är ett *tillräckligt* minimeringsvillkor av andra graden.

Begränsningar hos villkoren

Informationen att $\nabla f(x_*) = 0$ och $\nabla^2 f(x_*)$ är positivt semi-definit räcker inte till för att avgöra om en punkt är en minimerare.

Exempel:

$\mathbf{f(x)}$	$\mathbf{\nabla f(x)}$	$\mathbf{\nabla^2 f(x)}$	$\mathbf{\nabla^2 f(x_*)}$		$\mathbf{x_* = 0}$
x^2	$2x$	2	2	> 0	min
x^3	$3x^2$	$6x$	0	≥ 0	ej min, ej max
x^4	$4x^3$	$12x^2$	0	≥ 0	min
$-x^4$	$-4x^3$	$-12x^2$	0	≥ 0	max

Newtons klassiska metod för minimering

För att använda Newtons metod till att finna en minimerare applicerar vi 1:a ordningens nödvändiga villkor på en funktion f:

$$\nabla f(x) = 0 \qquad (f'(x) = 0)$$

Detta ger newtonsekvensen

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k) \qquad (x_{k+1} = x_k - f'(x)/f''(x))$$

Detta skrivs ofta som $x_{k+1} = x_k + p_k$, där p_k är lösningen till Newtons ekvation:

$$\nabla^2 f(x_k) p_k = -\nabla f(x_k),$$

vilket tydliggör att det i varje steg löses ett linjärt ekvationssystem (och inte beräknas en invers).

Newtons klassiska metod, forts.

Notera att approximationen av det icke-linjära funktionen $\nabla f(x)$ med

$$\nabla f(x_k + p) \approx \nabla f(x_k) + \nabla^2 f(x_k) p$$

motsvarar att approximera den icke-linjära funktionen $f(x)$ med den kvadratiska funktionen

$$Q(p) \equiv f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p,$$

dvs de tre första termerna i taylorutvecklingen av f runt x_k.

En användbar tolkning av Newtons metod blir att vi i varje iterationssteg approximerar f med en kvadratisk funktion Q och beräknar x_{k+1} som minimerare av Q.

Egenskaper hos Newtons metod

Fördel:

- Den konvergerar typiskt kvadratiskt mot en stationär punkt.

Nackdelar:

- Den konvergerar ej nödvändigtvis mot en minpunkt.
- Den kan divergera om startpunkten ligger för långt från lösningen.
- Den kan misslyckas om $\nabla^2 f(x_k)$ ej inverterbar för något k.
- Den behöver andraderivateinformation i $\nabla^2 f(x_k)$.
  - "Jobbigt" ta fram funktionerna. Kan bli fel.
  - Kräver beräkning och lagring av n^2 element.
  - Lösningen av Newtons ekvationer kräver n^3 operationer.

Newtons metod, forts.

I praktiken används Newtons metod i sin klassiska formulering aldrig. Däremot är den grunden för många andra algoritmer som söker minska dess negativa egenskaper på olika sätt och samtidigt behålla metodens positiva egenskaper (främst konvergenshastigheten).

Garantera nedåtriktning

Enligt vår generella optimeringsalgoritm bestäms den nya approximationen på formen

$$x + \alpha p, \quad \alpha > 0, \quad f(x + \alpha p) < f(x).$$

Detta är möjligt endast om p är en nedförsriktning, dvs

$$p^T \nabla f(x) < 0.$$

I Newtons metod bestäms sökriktningen som

$$p = -\nabla^2 f(x)^{-1} \nabla f(x).$$

Om p ska vara en nedåtriktning i punkten x måste

$$p^T \nabla f(x) = -\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) < 0$$

eller

$$\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) > 0$$

Detta uppfylls om $\nabla^2 f(x)^{-1}$ är positivt definit.

Garantera nedåtriktning, forts.

Kravet att $\nabla^2 f(x)$ ska vara positivt definit är hårdare än $p^T \nabla f(x) < 0$.

Newtons metod går att tolka som att funktionen $f(x+p)$ approximeras med en kvadratisk funktion

$$f(x+p) \approx Q(p) \equiv f(x) + p^T \nabla(f) + \frac{1}{2} p^T \nabla^2 f(x) p.$$

Newtons formel fås genom att lösa $\nabla Q(p) = 0$.

Med denna formel konvergerar Newtons metod mot en stationär punkt, men då vårt mål var att minimera f, väljer vi i stället att tolka Newtons formel som ett försök att lösa $\min_p Q(p)$. Om $\nabla^2 f(x)$ är positivt definit är lösningar ekvivalenta, men om $\nabla^2 f(x)$ är indefinit saknar $Q(p)$ ett minimum.

Garantera nedåtriktning, forts.

Om hessianen är indefinit kan vi modifiera den så att den blir positivt definit. Då är vi garanterade en nedåtriktning, och lösningen motsvarar $\min_p Q(p)$.

Detta kan verka godtyckligt, men gjort på rätt sätt kommer denna modifierade algoritm att konvergera. Dessutom är hessianen nära lösningspunkten typiskt positivt definit, vilket innebär att denna modifiering endast utförs långt från lösningspunkten.

Denna modifiering av hessianen kan utföras för minimal extra kostnad: Sökriktningen p beräknas som lösningen till

$$\nabla^2 f(x) p = -\nabla f(x)$$

och om $\nabla^2 f(x)$ är positivt definit kan matrisfaktoriseringen

$$\nabla^2 f(x) = LDL^T$$

användas, där diagonalelementen i D är positiva.

Garantera nedåtriktning, forts.

Är $\nabla^2 f(x)$ ej positivt definit kommer ett diagonalelement d_{ii} i D någon gång under faktoriseringen att uppfylla $d_{ii} \leq 0$. Skulle detta hända kan d_{ii} ersättas med ett lämpligt positivt tal (den modifierade matrisen får t.ex. inte bli singulär).

Detta är ekvivalent med att utföra faktoriseringen på en modifierad matris

$$\nabla^2 f(x) + E = LDL^T,$$

där E är en diagonalmatris.

Sedan används den modifierade faktoriseringen till att beräkna sökriktningen

$$(LDL^T)p = -\nabla f(x).$$

Algoritm för Newtons modifierade metod

Bestäm startgissning x_0 och konvergenstolerans ε.

Upprepa för $k = 0, 1, \ldots$

- Om $\|\nabla f(x)\| < \varepsilon$, avsluta.

- Beräkna den modifierade faktoriseringen av hessianen

  $$\nabla^2 f(x_k) + E = LDL^T$$

  och lös

  $$(LDL^T)p = -\nabla f(x_k)$$

  för sökriktningen p_k.

- Utför linjesökning för att bestämma den nya uppskattningen $x_{k+1} = x_k + \alpha_k p_k$.

Garantera konvergens

De externa tekniker som används för att garantera konvergens försöker hejda optimeringsmetoderna när de är på väg att divergera och försöker samtidigt undvika att störa när optimeringsmetoden fungerar effektivt.

Långt från lösningen kommer taylorutvecklingen att vara en dålig approximation av funktionen nära lösningen och dessa "globala strategier" är en viktigt del av algoritmen och förhindrar steg från lösningen eller t.o.m. divergens. Nära lösningen blir strategierna sällan aktiverade men finns med i bakgrunden som säkerhetsspärrar.

Garantera konvergens, forts.

Termen "globala strategier" används för att skilja på metoderna för att bestämma en ny approximation till lösningen och beräkningen av sökriktningen. Bestämningen av sökriktningen är baserad på en lokal approximation av funktionen, medan bestämningen av nästa approximation av lösningen konstrueras för att garantera *global konvergens*, dvs konvergens från varje punkt.

Vi kommer att diskutera två huvudtyper av globala strategier, *linjesökning* och *trust-region*-metoder.

Linjesökning

Låt x_k vara nuvarande approximation av en minimerare till f och låt p_k vara sökriktningen i x_k. Den nya uppskattningen definieras som

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

där steglängden α_k väljs så att

$$f(x_{k+1}) < f(x_k),$$

dvs varje steg minskar funktionsvärdet och tar oss närmare lösningen.

Linjesökning, exempel

Studera problemet

$$\min f(x_1, x_2) = 5x_1^2 + 7x_2^2 - 3x_1x_2.$$

För $x_k = [2,\ 3]^T$ och $p_k = [-5,\ -7]^T$ är $f(x_k) = 65$. För $\alpha_k = 1$ blir

$$f(x_k + \alpha_k p_k) = f(-3, -4) = 121 > f(x_k),$$

dvs $\alpha_k = 1$ är inte en acceptabel steglängd. För $\alpha = \frac{1}{2}$ blir

$$f(x_k + \alpha_k p_k) = f(-\frac{1}{2}, -\frac{1}{2}) = \frac{9}{4} < f(x_k),$$

dvs $\alpha_k = \frac{1}{2}$ är en acceptabel steglängd.

Linjesökning, forts.

Vi antar att p_k är en nedåtriktning vid punkten x_k, dvs

$$p_k^T \nabla f(x_k) < 0.$$

Då kommer $f(x_k + \alpha p_k) < f(x_k)$ åtminstone för något tillräckligt litet $\alpha > 0$. Vi kommer därför att anta att $\alpha_k > 0$.

Termen *linjesökning* används därför att sökningen efter nästa approximation x_{k+1} görs längs med linjen $y(\alpha) = x_k + \alpha p_k$.

Helst skulle vi vilja välja α_k som lösningen till

$$\min_{a>0} F(\alpha) \equiv f(x_k + \alpha p_k). \qquad (2)$$

Detta kallas *exakt linjesökning*, och är ett endimensionellt minimeringsproblem som generellt är för dyrt för att lösas exakt. I praktiken löses problemet (2) därför endast approximativt.

Linjesökning, forts.

Ett sätt att garantera konvergens är att göra några antaganden:

Sökriktningen p_k (a) är "tillräckligt nedåtlutande" och är (b) "gradientrelaterad".

Steglängden α_k (a) producerar en "tillräcklig minskning" i objektfunktionen f, och är (b) inte "för liten".

Krav på sökriktningen

Villkoret "tillräckligt nedåtlutande" svarar mot att $p_k^T \nabla f(x)$ inte får bli hur nära 0 som helst, utan måste uppfylla

$$-\frac{p_k^T \nabla f(x_k)}{\|p_k\| \cdot \|\nabla f(x_k)\|} \geq \varepsilon > 0,$$

för alla k och något $\varepsilon > 0$.

Detta villkor kan också skrivas som

$$\cos \theta \geq \varepsilon > 0,$$

där θ är vinkeln mellan sökriktningen p_k och den negativa gradienten $-\nabla f(x_k)$, och refereras därför till som *vinkelvillkoret*.

Detta villkor innebär att vinkeln mellan sökriktningen och den negativa gradienten inte för bli hur nära 90 grader (ortogonalitet) som helst.

Krav på sökriktningen, forts.

Sökriktningen p_k sägs vara *gradientrelaterad* om

$$\|p_k\| \geq m \|\nabla f(x_k)\|$$

för alla k och något $m > 0$.

Detta villkor innebär att sökriktningen inte får bli hur liten som helst relativt gradienten.

Krav på steglängden

De två villkoren på sökriktningen kan normalt uppfyllas genom små förändringar i algoritmerna och antas tills vidare vara uppfyllt.

Att steglängden ska ge en "tillräcklig minskning" av objektfunktionen mäts med hjälp av den linjära taylorapproximationen

$$f(x_k + \alpha p_k) \approx f(x_k) + \alpha p_k^T \nabla f(x_k).$$

Vi kommer att begära att minskningen av objektfunktionen är en fraktion av den som predikteras av den linjära taylorapproximationen, eller

$$f(x_k + \alpha_k p_k) \leq f(x_k) + \mu \alpha_k p_k^T \nabla f(x_k),$$

där $0 < \mu < 1$. Detta kallas ibland för *Armijos villkor*.

Krav på steglängden, forts.

Ett sätt att hantera kravet att steglängden inte får vara "för liten" kallas *backtracking*.

Antag att p_k uppfyller "tillräckligt nedåtriktad"-villkoret. Definiera α_k som det första elementet i sekvensen

$$1, \frac{1}{2}, \frac{1}{4}, \ldots, 2^{-i}, \ldots$$

som uppfyller kravet på tillräcklig minskning av objektfunktionen. Ett sådant α_k finns alltid.

Eftersom stora steg testas före små ger denna algoritm inte "för små" steglängder.

Andra linjesökningar

En annan approximation till lösning på

$$\min_{\alpha>0} F(\alpha) \equiv f(x_k + \alpha p_k)$$

är att söka $F'(\alpha) = 0$, vilket approximeras till villkoret

$$|F'(\alpha_k)| \le \eta |F'(0)|, \qquad (3)$$

där η är en konstant $0 \le \eta < 1$.

Det går att visa att $F'(\alpha) = p_k^T \nabla f(x_k + \alpha p_k)$, vilket insatt i (3) ger

$$|p_k^T \nabla f(x_k + \alpha_k p_k)| \le \eta |p_k^T \nabla f(x_k)|.$$

och kallas för *Wolfes villkor*.

Observera att detta villkor också uppfylls av en maximerare till (3), varför Wolfes villkor måste kombineras med Armijos villkor för att säkerställa global konvergens.

Newtons metod, garanterad konvergens

För att en sökriktning ska ge garanterad konvergens, måste två villkor uppfyllas: Sökriktningen måste vara "tillräcklig nedåtriktad" (*sufficient descent*), dvs

$$-\frac{p_k^T \nabla f(x_k)}{\|p_k\| \|\nabla f(x_k)\|} \ge \epsilon > 0$$

för alla k och en specificerad konstant $\epsilon > 0$.

Dessutom ska sökriktningen vara "gradient-relativ" (*gradient related*), dvs

$$\frac{\|p_k\|}{\|\nabla f(x_k)\|} \ge m > 0$$

för alla k och en någon konstant $m > 0$.

Newtons metod, sufficient descent

Låt g beteckna $\nabla f(x_k)$ och H beteckna $\nabla^2 f(x_k)$. Låt $\|\cdot\| = \|\cdot\|_2$.

För Newtons metod är

$$p_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k) = -H^{-1} g,$$

så *sufficient descent*-villkoret blir:

$$\begin{aligned} -\frac{p_k^T \nabla f(x_k)}{\|p_k\| \|\nabla f(x_k)\|} &= -\frac{-g^T H^{-1} g}{\|H^{-1} g\| \|g\|} \\ &= \frac{g^T H^{-1} g}{\|H^{-1} g\| \|g\|} \end{aligned}$$

Om H är positivt definit, så har H positiva egenvärden, dvs det existerar en faktorisering $H = V \Lambda V^T$, där Λ är en diagonalmatris med positiva diagonalelement och $V^T V = I$. Då blir $H^{-1} = V \Lambda^{-1} V^T$. Låt dessutom $\hat{g} = g / \|g\|$.

Newtons metod, sufficient descent, forts.

Insättning ger:

$$\begin{aligned} \frac{g^T H^{-1} g}{\|H^{-1} g\| \|g\|} &= \frac{\hat{g}^T V \Lambda^{-1} V^T \hat{g} \|g\|^2}{\|V \Lambda^{-1} V^T \hat{g}\| \|g\|^2} \\ &= [u = V^T \hat{g}] \\ &= \frac{u^T \Lambda^{-1} u}{\|V \Lambda^{-1} u\|} \\ &= \frac{u^T \Lambda^{-1} u}{\|\Lambda^{-1} u\|} \\ &\ge \frac{1/\lambda_{max}}{1/\lambda_{min}} \\ &= \frac{\lambda_{min}}{\lambda_{max}}, \end{aligned}$$

där $\lambda_{max} = \max \lambda_{ii}$ och $\lambda_{min} = \min \lambda_{ii}$ är största resp. minsta egenvärdet hos H.

Om $H = \nabla^2 f(x_k)$ är positivt definit för alla k och $\frac{\ell_{min}}{\ell_{max}} = \frac{\min_k \lambda_{min}}{\max_k \lambda_{max}} = \epsilon > 0$, så är newtonriktningen garanterad att ge tillräcklig minskning.

Newtons metod, gradientrelativ

Med samma resonemang och substitutioner:

$$\begin{aligned}\frac{\|p_k\|}{\|\nabla f(x_k)\|} &= \frac{\|H^{-1}g\|}{\|g\|} \\ &= \frac{\|V\Lambda^{-1}V^T\hat{g}\|\|g\|}{\|g\|} \\ &= \|V\Lambda^{-1}u\| \\ &= \|\Lambda^{-1}u\| \\ &\geq \frac{1}{\lambda_{max}}.\end{aligned}$$

Om $H = \nabla^2 f(x_k)$ är positivt definit för alla k och $\frac{1}{\ell_{max}} = \frac{1}{\max_k \lambda_{max}} = m > 0$, så är newtonriktningen gradientrelativ.

Trust-region-metoder

Trust-region-metoder refererar explicit till en "modell" av objektfunktionen, i Newtons fall den kvadratiska modellen härledd ur taylorapproximationen av f runt punkten x_k:

$$\psi_k(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p$$

Denna modell "litas på" endast inom ett begränsat område runt x_k, definierat genom

$$\|p\| \leq \Delta_k.$$

Detta kommer att begränsa längden på steget från x_k till x_{k+1}.

Värdet på Δ_k kommer att justeras upp om modellen överensstämmer "väl" med objektfunktionen, och ner om den stämmer "dåligt".

Trust-region-metoder, forts.

Vid iteration k av en trust-region-metod löses följande delproblem:

$$\begin{aligned}\min_p \psi_k(p) &= f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p \\ &\text{s.t.} \qquad\qquad\qquad\qquad\qquad (4) \\ \|p\| &\leq \Delta_k\end{aligned}$$

Lösningen till detta bivillkorsproblem är lösningen till det linjära ekvationssystemet

$$(\nabla^2 f(x_k) + \lambda I) p_k = -\nabla f(x_k),$$

där $\lambda \geq 0$, matrisen $(\nabla^2 f(x_k) + \lambda I)$ är positivt definit och $\lambda(\Delta_k - \|p_k\|) = 0$.

Trust-region-metoder, forts.

Notera att om $\nabla^2 f(x_k)$ är positivt definit och Δ_k tillräckligt stor är lösningen till (4) lösningen till

$$\nabla^2 f(x_k) p = -\nabla f(x_k),$$

dvs p är en *newtonriktning*.

I övrigt är

$$\Delta_k \geq \|p_k\| = \|(\nabla^2 f(x_k) + \lambda I)^{-1} \nabla f(x_k)\|,$$

så om $\Delta_k \to 0$ så $\lambda \to \infty$ och

$$p_k \to -\frac{1}{\lambda} \nabla f(x_k).$$

När λ varierar mellan 0 och ∞ kan man visa att $p_k(\lambda)$ varierar mellan newtonriktningen och en multipel av den negativa gradienten.

Detta skiljer sig från linjesökningen där sökriktningen beräknas först och sedan hålls fix då steglängden beräknas.

Trust-region-algoritm

Bestäm startgissning x_0, initial trust-regiongräns Δ_0 och konstanter $0 < \mu < \eta < 1$.

Upprepa för $k = 0, 1, \ldots$

- Om x_k är optimal, avsluta.

- Lös

$$\begin{aligned} \min_p \psi_k(p) &= f(x_k)+\nabla f(x_k)^T p+\tfrac{1}{2}p^T\nabla^2 f(x_k)p \\ &\text{s.t.} \\ \|p\| &\leq \Delta_k \end{aligned}$$

för ett försökssteg p_k.

- Beräkna

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{f(x_k) - \psi(p_k)} = \frac{\text{verklig reduktion}}{\text{predikterad reduktion}}.$$

Trust-region-algoritm, forts.

- Om $\rho_k \leq \mu$
  - $x_{k+1} = x_k$ (misslyckat steg)

  annars
  - $x_{k+1} = x_k + p_k$ (lyckat steg).

- Uppdatera Δ_k:

$$\begin{aligned} \rho_k \leq \mu \Rightarrow &\quad \Delta_{k+1} = \tfrac{1}{2}\Delta_k \\ \mu < \rho_k < \eta \Rightarrow &\quad \Delta_{k+1} = \Delta_k \\ \rho_k \geq \eta \Rightarrow &\quad \Delta_{k+1} = 2\Delta_k \end{aligned}$$

Dogleg-algoritmen

Trust-region-delproblemet

$$\begin{aligned} \min_p \psi_k(p) &= f(x_k) + \nabla f(x_k)^T p + \frac{1}{2}p^T\nabla^2 f(x_k)p \\ &\text{s.t.} \\ \|p\| &\leq \Delta_k \end{aligned}$$

är besvärligt. Om newtonriktningen inte uppfyller bivillkoret och vi vill ta ett så långt steg som möjligt, måste vi bestämma λ så att

$$\|p_k\| = \|(\nabla^2 f(x_k) + \lambda I)^{-1}\nabla f(x_k)\| = \Delta_k,$$

vilket är en icke-linjär ekvation i λ.

Ett sätt är att lösa detta approximativt är att approximera funktionen $p_k(\lambda)$ med styckvis linjärt för att finna ett λ så att $\|p_k(\lambda)\| = \Delta_k$. Denna metod kallas "dogleg" efter en golfterm.

En annan lösning är att beräkna $p_k(\lambda)$ för $\lambda_i = 0, \lambda_{min}, k\lambda_{min}, \ldots$ och välja det första $\|p_k(\lambda_i)\| \leq \Delta_k$.

Överväganden

Om vi lägger ner stor möda (programmerings- och/eller beräkningstid) på att lösa (4), dvs ta ett så bra steg som möjligt kommer vi *förmodligen* att behöva färre Newton-iterationer.

Vilken metod som i slutänden är snabbast blir alltså en avvägning mellan bra, tidskrävande, men få steg och dåliga, snabba, men många steg.

Detta är en mycket vanligt problem.