Aalto-yliopisto
Perustieteiden
korkeakoulu

Epälineaaristen pienimmän neliösumman tehtävien ratkaiseminen numeerisilla optimointimenetelmillä (valmiin työn esittely)

Lari Pelkola

29.9.2014

Ohjaaja/valvoja: Prof. Harri Ehtamo

Systeemianalyysin laboratorio

Esityksen sisältö

- Tausta
- Tavoitteet
- Epälineaariset PNS-tehtävät
- Menetelmät
  - Kvasi-Newton (BFGS)
  - Gauss-Newton
  - Levenberg-Marquardt
- Tulokset
- Yhteenveto
- Tietolähteet

Tausta

- Parametrien estimointi
  - Mittausdata ja matemaattinen malli
    - Mallin parametrit pyrkivät selittämään esimerkiksi fysikaalisia ominaisuuksia mahdollisimman hyvin
- Optimaalisten parametrien löytämiseksi muodostetaan optimointimalli
  - Pienimmän neliösumman tehtävä

Tavoitteet

- Tutustua epälineaarisiin PNS-tehtäviin yleisellä tasolla
- Esitellä kolme numeerista ratkaisumenetelmää
  - kvasi-Newton
  - Gauss-Newton
  - Levenberg-Marquardt
- Testata algoritmeja muutamalla esimerkkitehtävällä
  - Iteraatioiden lukumäärät
  - suppeneminen
  - alkuarvauksen vaikutus

Epälineaariset PNS-tehtävät

- Yleinen muoto

$$\min_{x \in \Re^n} f(x) = \|r(x)\|^2 = \sum_{i=1}^{m} (r_i(x))^2 = \sum_{i=1}^{m} (a(t_i, x) - y_i)^2$$

- m epälineaarinen x:n suhteen, ei voida ratkaista analyyttisesti
  - numeeriset menetelmät
- Derivaatat:

$$\nabla f(x) = 2J(x)^T r(x)$$

$$H(x) = 2J(x)^T J(x) + 2S(x), \quad S(x) = \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x)$$

  - J(x) on r(x):n Jacobin matriisi

Epälineaariset PNS-tehtävät

- Ratkaisemiseen kaksi lähestymistapaa
- 1) Seuraa epälineaaristen yhtälöryhmien ratkaisemisesta r(x) = 0.
  - residuaalifunktion r(x) lineaarinen approksimaatio:

$$\min_x \left\| r(x_k) + J(x_k)(x - x_k) \right\|^2$$

  - Gauss-Newton ja Levenberg-Marquardt
- 2) Rajoittamattoman minimoinnin erikoistapaus
  - kohdefunktion f(x) kvadraattinen approksimaatio:

$$f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T H(x_k)(x - x_k)$$

Menetelmät

- Kvasi-Newton BFGS-menetelmä
  - Paljon käytetty menetelmä yleisesti epälineaarisessa optimoinnissa
  - Sopii tilanteisiin, joissa Hessen matriisi raskas laskea
  - Hessen matriisille muodostetaan arvio B_k , jota arvioidaan iteratiivisesti gradienttien avulla
  - B_k:t pysyvät positiivisesti definiitteinä iteraation kuluessa
  - Ei ota huomioon PNS-tehtävän Hessen matriisin rakennetta, vaan approksimoi suoraan koko Hessen matriisia
  - Iteraatioaskeleen pituus λ_k saadaan viivahaulla
  - Iteraatioaskel: $x_{k+1} = x_k - \lambda_k B_k^{-1} \nabla f(x_k)$

Menetelmät

- Gauss-Newton
  - yksinkertaisin vaihtoehto epälineaarisiin PNS-tehtäviin
  - iteraatioaskel aiemmin esitetyn lineaarisen PNS-tehtävän ratkaisu: $x_{k+1} = x_k - (J(x_k)^T J(x_k))^{-1} J(x_k)^T r(x_k)$
  - Eroaa Newtonin iteraatiosta vain siten, että Hessen matriisin toisen kertaluvun informaatio S(x) jätetään huomioitta
    - Saattaa hajaantua suuriresiduaalisissa tai vahvasti epälineaarisissa tehtävissä, eli kun S(x) on suuri
  - Jos J(x) on esimerkiksi häiriöaltis, menetelmä voi törmätä singulariteetteihin
  - Suppeneminen voidaan saada varmemmaksi viivahaulla

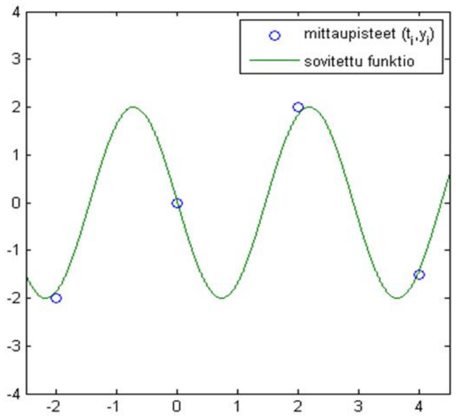
Menetelmät

- Levenberg-Marquardt
  - Käytetyimpiä menetelmiä epälineaarisissa PNS-tehtävissä
  - Käyttää luottamusalueeseen liittyvää logiikkaa suppenemisen varmistamiseksi
  - Iteraatioaskel: $x_{k+1} = x_k - (J(x_k)^T J(x_k) + \mu_k I)^{-1} J(x_k)^T r(x_k)$, tai

    $x_{k+1} = x_k - (J(x_k)^T J(x_k) + \mu_k diag(J(x_k)^T J(x_k))^{-1} J(x_k)^T r(x_k)$
  - Parametri $\mu_k \geq 0$ ohjaa askeleen suuntaa ja pituutta:
    - Jos μ_k on pieni, on menetelmä lähellä Gauss-Newton askelta
    - Jos μ_k on suuri, on menetelmä lähellä gradienttimenetelmää, jolloin myös askeleen pituus on pieni

Tulokset

- Yksinkertainen sinifunktio: $a(t,x) = 2\sin(x_1 t + x_2)$



(t1,y1)	(t2,y2)	(t3,y3)	(t4,y4)
(-2,-2)	(0,0)	(2,2)	(4,-1.5)

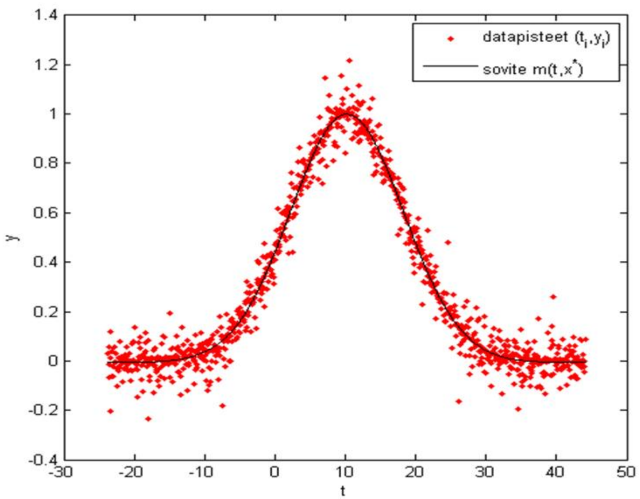
	BFGS	LM	LMd	GN	GNv
Iter.lkm	9	7	7	10	6

- Suuremman residuaalin vaikutus: vaihdetaan y3 = 6

	BFGS	LM	LMd	GN	GNv
Iter.lkm	9	12	12	*	18

Tulokset

- Gaussin funktio: $\mathrm{a}(\mathrm{t},\mathrm{x}) = x_1 e^{\frac{-(t-x_2)^2}{2x_3^{\,2}}} + x_4$



$$x_0 = (1,10,10,0)^T$$

	BFGS	LM	LMd	GN	GNv
Iter.lkm	25	4	4	8	8

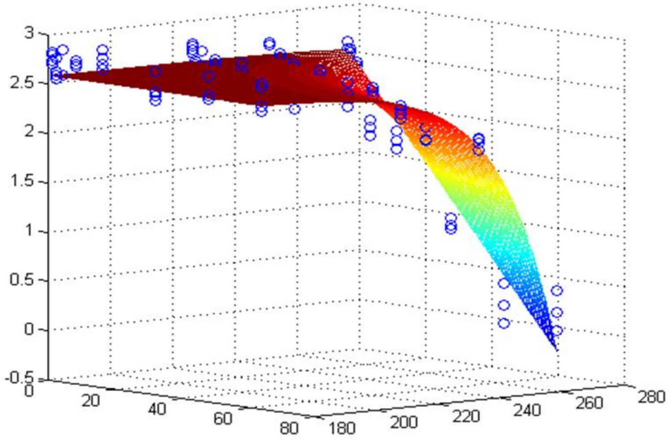
Huonolla alkuarvauksella:

$$x_0 = (100,-20,40,-20)^T$$

	BFGS	LM	LMd	GN	GNv
Iter.lkm	17, ei globaali optimi	21	27	*	*

Kolmiulotteinen esimerkki

- Malli: $a(t_1, t_2, x) = x_1 + x_2 t_1 e^{-x_3 t_2}$



$$x_0 = (2.5, 5 \cdot 10^{-9}, -5 \cdot 10^{-2})^T$$

	BFGS	LM	LMd	GN	GNv
Iter.lkm	-	95	32	*	41

$$x_0 = (2, 10^{-4}, -10^{-2})^T$$

	BFGS	LM	LMd	GN	GNv
Iter.lkm	132	229, ei globaali optimi	162	*	40

Yhteenveto

- Yleispäteviä johtopäätöksiä menetelmien suorituksista vaikea tehdä, tehtävien rakenne ja alkuarvaus vaikuttavat algoritmien suorituksiin lähes ainutlaatuisella tavalla
- BFGS suhteellisen robusti ja yleispätevä menetelmä, näissä esimerkeissa usein hitain
- GN toimii parhaiten pienen residuaalin tehtävissä ja hyvällä alkuarvauksella
  - voi hajaantua vahvasti epälineaarisissa tehtävissä tai jos alkuarvaus huono
  - suppenemisen saa varmemmaksi viivahaulla
- LM ehkä näistä kolmesta soveltuvin, robustimpi kuin GN

Tietolähteet

- J. Haataja: Optimointitehtävien ratkaiseminen (CSC – Tieteellinen laskenta Oy, 1995, 2.painos. Suurten optimointitehtävien ratkaiseminen, CSC Tieteellinen laskenta Oy, 1994
- J.E Dennis, JR ja R.B Schnabel: Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice Hall Series in Computational Mathematics, 1983
- Å. Björck: Numerical Methods for Least Squares Problems, Society for Industrial and Applied Mathematics, 1996

Aineisto

- Gaussin funktion testidata:
  - Bucknell university, Department of Physics and Astronomy,
  - Models and fitting
- Kolmiulotteisen esimerkin testidata:
  - National Institute of Standards and Technology, StrD Nonlinear Least Squares Regression Datasets