Pienimmän neliösumman tehtävät

Tarkastellaan menetelmiä pienimmän neliösumman tehtävien ratkaisemiseen. Tällaiset tehtävät ovat muotoa

$$
\begin{aligned}
(1) \qquad & \text{minimoi} \quad F(\mathbf{x}) = \frac{1}{2}\|\mathbf{f}(\mathbf{x})\|^2 = \frac{1}{2}\sum_{i=1}^{m}(f_i(\mathbf{x}))^2 \\
& \text{rajoittein } \ \mathbf{x} \in \mathbf{R}^n.
\end{aligned}
$$

Ne ovat käytännössä hyvin yleisiä. Ensisijainen käyttö on silloin, kun $\mathbf{f}$ koostuu n:stä reaaliarvoisesta funktiosta ja halutaan ratkaista n yhtälön ryhmä, jossa on n muuttujaa $\mathbf{f}(\mathbf{x}) = \mathbf{0}$. Se voidaan muotoilla pienimmän neliösumman tehtäväksi muotoon (1).

Tämä siksi, että $\mathbf{x}^*$ on yhtälöryhmän $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ ratkaisu jos ja vain jos se minimoi funktion $\frac{1}{2}\|\mathbf{f}(\mathbf{x})\|^2$ ja optimiarvo on nolla.

Käyrän sovitus

Oletetaan, että halutaan arvioida matetaattisen mallin n parametria mittaustiedon avulla siten, että malli esittää mahdollisimman hyvin fysikaalista systeemiä.

Oletetaan, että meillä on approksimoiva relaatio muotoa

$$z = h(x, y),$$

missä h on tunnettu mallia kuvaava funktio ja

$x \in \mathbf{R}^n$ on tuntemattomien parametrien vektori
$z \in \mathbf{R}^r$ on mallin tulos
$y \in \mathbf{R}^p$ on mallin syöte.

Jos olemme fysikaalisista mittauksista saaneet m syöte-tulos -paria $(y_1, z_1), \ldots, (y_m, z_m)$ mallitettavasta systeemistä, haluamme löytää sellaisen parametrivektorin x, joka mahdollisimman hyvin sopii dataan. Tätä mitataan siten, että virheiden neliöiden summa on mahdollisimman pieni. Siis minimoidaan funktioita

$$\frac{1}{2} \sum_{i=1}^{m} (z_i - h(x, y_i))^2.$$

Jos esimerkiksi halutaan sovittaa tietoparit kuutiollisisella polynomiapproksimaatiolla, valitaan

$$h(x, y) = x_3 y^3 + x_2 y^2 + x_1 y + x_0,$$

missä $x = (x_0, x_1, x_2, x_3)$ on kuutiollisen polynomin tuntemattomien kerrointen vektori.

Muita pienimmän neliösumman tehtävän sovelluksia löytyy mm. dynaamisten systeemien identifioinnista, hermoverkoista jne.

Tarkastellaan seuraavaksi erikoismenetelmiä pienimmän neliösumman tehtävän ratkaisemiseen. Minimoidaan siis funktiota $\frac{1}{2}\|\mathbf{f}(\mathbf{x})\|^2$.

Gaussin ja Newtonin menetelmä

Yleisin pienimmän neliösumman tehtävän ratkaisumenetelmä on Gaussin ja Newtonin menetelmä. Olkoon meillä piste $\mathbf{x}^h$. Menetelmän perusmuoto perustuu funktion $\mathbf{f}$ linearisointiin, jolloin saadaan

$$\hat{\mathbf{f}}(\mathbf{x}, \mathbf{x}^h) = \mathbf{f}(\mathbf{x}^h) + \nabla \mathbf{f}(\mathbf{x}^h)^T (\mathbf{x} - \mathbf{x}^h).$$

Linearisoidun funktion $\hat{\mathbf{f}}$ normin minimoimiseksi saadaan

(2)

$$\begin{aligned}\mathbf{x}^{h+1} =& \arg\min_{\mathbf{x}\in\mathbf{R}^n} \frac{1}{2}\|\hat{\mathbf{f}}(\mathbf{x},\mathbf{x}^h)\|^2 = \arg\min_{\mathbf{x}\in\mathbf{R}^n} \frac{1}{2}\{\|\mathbf{f}(\mathbf{x}^h)\|^2 \\ &+ 2(\mathbf{x}-\mathbf{x}^h)^T\nabla\mathbf{f}(\mathbf{x}^h)\mathbf{f}(\mathbf{x}^h) \\ &+ (\mathbf{x}-\mathbf{x}^h)^T\nabla\mathbf{f}(\mathbf{x}^h)\nabla\mathbf{f}(\mathbf{x}^h)^T(\mathbf{x}-\mathbf{x}^h)\},\end{aligned}$$

missä $\nabla\mathbf{f}(\mathbf{x}^h)$ on vektoriarvoisen funktion $\mathbf{f}$ Jacobin matriisi pisteessä $\mathbf{x}^h$.

Olettaen, että $n \times n$-matriisi $\nabla\mathbf{f}(\mathbf{x}^h)\nabla\mathbf{f}(\mathbf{x}^h)^T$ on kääntyvä, kvadraattinen minimointi yllä tuottaa

(3)

$$x^{h+1} = \mathbf{x}^h - (\nabla\mathbf{f}(\mathbf{x}^h)\nabla\mathbf{f}(\mathbf{x}^h)^T)^{-1}\nabla\mathbf{f}(\mathbf{x}^h)\mathbf{f}(\mathbf{x}^h).$$

Huomattakoon, että jos $\mathbf{f}$ on lineaarinen, niin

$$\|\mathbf{f}(\mathbf{x}^h)\|^2 = \|\hat{\mathbf{f}}(\mathbf{x},\mathbf{x}^h)\|^2$$

ja menetelmä konvergoi yhdellä iteraatiolla.

Huomaa myös, että hakusuunta on laskeva, sillä $\nabla\mathbf{f}(\mathbf{x}^h)\mathbf{f}(\mathbf{x}^h)$ on pienimmän neliösumman funktion $\frac{1}{2}\|\mathbf{f}(\mathbf{x})\|^2$ gradientti pisteessä $\mathbf{x}^h$ ja $(\nabla\mathbf{f}(\mathbf{x}^h)\nabla\mathbf{f}(\mathbf{x}^h)^T)^{-1}$ on positiivisesti definiitti.

Gaussin ja Newtonin menetelmän muunnoksia

Menetelmää voidaan muokata hakusuunnan laskevuuden takaamiseksi ja jotta kyettäisiin käsittelemään tilanteita, joissa matriisi $\nabla \mathbf{f}(\mathbf{x}^h)\nabla \mathbf{f}(\mathbf{x}^h)^T$ on singulaarinen. Muokattua menetelmää voidaan käyttää myös konvergenssin vauhdittamiseksi silloin, kun em. matriisi on lähes singulaarinen.

Gaussin ja Newtonin menetelmä voidaan muuntaa muotoon

$$\mathbf{x}^{h+1} = \mathbf{x}^h - \lambda^h (\nabla \mathbf{f}(\mathbf{x}^h)\nabla \mathbf{f}(\mathbf{x}^h)^T + \Delta^h)^{-1} \nabla \mathbf{f}(\mathbf{x}^h)\mathbf{f}(\mathbf{x}^h),$$

missä λ^h on optimaalinen askelpituus hakusuuntaan ja Δ^h on diagonaalimatriisi siten, että matriisi

$$\nabla \mathbf{f}(\mathbf{x}^h)\nabla \mathbf{f}(\mathbf{x}^h)^T + \Delta^h$$

on positiivisesti definiitti.

Levenbergin ja Marquardtin menetelmä

Yksinkertainen tapa matriisin Δ^h valintaan on identtisen matriisin positiivinen monikerta. Tällöin menetelmän hakusuunnat ovat gradienttisukuisia.

Yhteys Newtonin menetelmään

Gaussin ja Newtonin menetelmä on läheistä sukua Newtonin menetelmälle. Pienimmän neliösumman funktion $\frac{1}{2}\|\mathbf{f}(\mathbf{x})\|^2$ Hessen matriisi on muotoa

$$\nabla \mathbf{f}(\mathbf{x}^h)\nabla \mathbf{f}(\mathbf{x}^h)^T + \sum_{i=1}^{m} \nabla^2 f_i(\mathbf{x}^h) f_i(\mathbf{x}^h).$$

Voidaan siis nähdä, että Gaussin ja Newtonin iteraatiot (2) ja (3) approksimoivat Newtonin menetelmää, kun toisen asteen termi

$$\sum_{i=1}^{m} \nabla^2 f_i(\mathbf{x}^h) f_i(\mathbf{x}^h)$$

jätetään pois.

Tällä tavoin Gaussin ja Newtonin menetelmässä säästetään laskennassa. Maksettava hinta on konvergenssin hidastuminen jonkin verran. Kuitenkin, jos pois jätetty termi on suhteellisen pieni optimin lähellä, konvergenssin aste on tyydyttävä Gaussin ja Newtonin menetelmässä. Tämä pätee useissa sovelluksissa, esimerkiksi silloin, kun $\mathbf{f}$ on miltei lineaarinen ja kun komponentit $f_i(\mathbf{x})$ ovat pieniä optimin läheisyydessä.

Jos $m = n$ ja ratkottavana on yhtälöryhmä

$$\mathbf{f}(\mathbf{x}) = \mathbf{0},$$

hylätty termi on optimissa nolla. Tällöin (olettaen, että $\nabla \mathbf{f}(\mathbf{x}^h)$ on käännettävissä) saamme

$$\begin{aligned}&(\nabla \mathbf{f}(\mathbf{x}^h) \nabla \mathbf{f}(\mathbf{x}^h)^T)^{-1} \nabla \mathbf{f}(\mathbf{x}^h) \mathbf{f}(\mathbf{x}^h)\\ &= (\nabla \mathbf{f}(\mathbf{x}^h)^T)^{-1} \mathbf{f}(\mathbf{x}^h)\end{aligned}$$

ja Gaussin ja Newtonin menetelmä alkuperäisessä asussaan saa muodon

$$\mathbf{x}^{h+1} = \mathbf{x}^h - (\nabla \mathbf{f}(\mathbf{x}^h)^T)^{-1} \mathbf{f}(\mathbf{x}^h),$$

mikä on täsmälleen sama kuin Newtonin menetelmä yhtälöryhmän $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ ratkaisemiseksi. Tällaisessa tilanteessa konvergenssi on superlineaarista.