Inversio-ongelmien laskennallinen peruskurssi
Luento 4

Kevät 2012

1 Regularisointi

Eräs keino yrittää ratkaista (likimääräisesti) huonosti asetettuja ongelmia on regularisaatio. Regularisoinnissa ongelmaa muutetaan niin, että se muuttuu hyvin asetetuksi (tai ainakin vähemmän huonosti asetetuksi). Tällöin tietenkin myös saatu ratkaisu muuttuu, eli liiallisen regularisoinnin vaarana on, että halutun approksimatiivisen ratkaisun sijasta saadaankin täysin väärä ratkaisu.

Erilaisia regularisointimenetelmiä on lukuisia, mutta tällä kurssilla tutustutaan lähemmin vain kahteen: typistettyyn singulaariarvohajotelmaan ja Tikhonovin regularisaatioon.

2 Typistettu singulaariarvohajotelma (Truncated SVD)

Oletetaan, että $A \in \mathbb{R}^{m \times n}$. Tarkastellaan lineaarisen yhtälön

$$m = Ax$$

ratkaisemista singulaariarvohajotelmaa $A = U\Sigma V^T$ käyttäen. Nyt matriisi $\Sigma\mathbb{R}^{m\times n}$ on diagonaalimatriisi, jonka diagonaali Σ_{diag} on muotoa

$$\Sigma_{diag} = (\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0),$$

missä σ_r on pienin nollasta eroava matriisin A singulaariarvo, $r \leq \min(m, n)$. Miniminormi (tai pienimmän neliösumman) ratkaisu voidaan kirjoittaa muodossa

$$\tilde{x} = V * \Sigma^+ * U^T = \sum_{i=1}^{r} \frac{1}{\sigma_i} \langle m, u_i \rangle v_i,$$

missä $\langle\cdot,\cdot\rangle$ on normaali vektorien sisätulo, u_i on matriisin U i:s sarakevektori ja v_i on matriisin V i:s sarakevektori.

Likimääräisen ratkaisun TSVD-estimaatti $\hat{x}^{(k)}$, $k \leq r$, annetaan kaavalla

$$\hat{x}^{(k)} = \sum_{i=1}^{k} \frac{1}{\sigma_i} \langle m, u_i \rangle v_i.$$

Toisin sanoen TSVD-estimaattiin otetaan mukaan vain k ensimmäistä singulaariarvoa ja niitä vastaavat matriisien U ja V sarakkeet. Tämä estimaatin antama ratkaisu $\hat{x}^{(k)}$ luonnollisesti eroaa likimääräisestä ratkaisusta $\tilde{x}$. Huomaa, että jos matriisilla A on 0-singulaariarvoja, niin likimääräinen ratkaisu $\tilde{x}$ on itsessään TSVD-estimaatti:

$$\tilde{x} = \hat{x}^{(r)}.$$

Tarkastellan nyt ongelmaa, joka sisältää virhetermin ε.

$$m = Ax + \varepsilon.$$

Ongelman (tarkan) ratkaisun x ja sen TSVD-estimaatin $\hat{x}^{(k)}$ eron selvittämiseksi kirjoitetaan ensin ratkaisu x käyttämällä kantana matriisin V sarakevektoreita (Huomaa, että koska V on ortogonaalinen, muodostavat sen sarakevektorit kannan avaruudessa $\mathbb{R}^n$.):

$$x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i = \sum_{i=1}^{n} x_{v_i} v_j.$$

Nyt

$$\hat{x}^{(k)} = \sum_{i=1}^{k} \frac{1}{\sigma_i} \langle m, u_i \rangle v_i = \sum_{i=i}^{k} \frac{1}{\sigma_i} \langle Ax, u_i \rangle v_i + \sum_{i=1}^{k} \frac{1}{\sigma_i} \langle \varepsilon, u_i \rangle v_i,$$

mutta

$$\langle Ax, u_i \rangle = \langle \sum_{l=1}^{n} \sigma_l \langle x, v_l \rangle u_l, u_i \rangle = \sigma_i \langle x, v_i \rangle = \sigma_i x_{v_i},$$

koska matriisin U ortogonaalisuuden perusteella $\langle u_i, u_j \rangle = 0$, kun $i \neq j$. Täten siis saadaan

$$\hat{x}^{(k)} = \sum_{i=1}^{k} x_{v_i} v_i + \sum_{i=1}^{k} \frac{1}{\sigma_i} \langle m, u_i \rangle v_i.$$

Nyt ratkaisun x ja sen TSVD-estimaatin $\hat{x}^{(k)}$ erolle saadaan

$$x - \hat{x}^{(k)} = \sum_{i=k+1}^{n} x_{v_i} v_i + \sum_{i=1}^{k} \frac{1}{\sigma_i} \langle m, u_i \rangle v_i,$$

ja

$$
\begin{aligned}
||x - \hat{x}^{(k)}||^2 &= ||\sum_{i=k+1}^{n} x_{v_i} v_i||^2 + ||\sum_{i=1}^{k} \frac{1}{\sigma_i} \langle \varepsilon, u_i \rangle v_i||^2 \\
&= \sum_{i=k+1}^{n} x_{v_i}^2 + \sum_{i=1}^{k} \frac{1}{\sigma_i^2} |\langle \varepsilon, u_i \rangle|^2.
\end{aligned}
$$

Ensimmäinen termi in riippumaton virheestä ε, ja sitä kutsutaan usein estimaatin $\hat{x}^{(k)}$ harhaksi (bias). Sille pätee

$$
\sum_{i=k+1}^{n} x_{v_i}^2 \to 0, \quad \text{kun } k \to n.
$$

Toista termiä kutsutaan rekonstruktiovirheeksi ja se kasvaa, kun $k \to n$. Varsinkin, jos matriisin A pienimmät singulaariarvot ovat hyvin pieniä (tai nollia), se saattaa kasvaa kontrolloimattomasti. Typistettyä singulaariarvohajotelmaa käytettäessä onkin tärkeää valita typistysparametri k niin, että

1. k on tarpeeksi suuri, jotta bias ei ole liian suuri, ja toisaalta
2. k on riittävän pieni, jotta rekonstruktiovirhe ei pääse kasvamaan liian suureksi.

Ideaalinen valinta olisi tietysti $k = k_{opt}$, siten, että

$$
k_{opt} = \min_k ||x - \hat{x}^{(k)}||^2.
$$

Tietenkään normaalisti meillä ei ole tietoa, mikä x on, vaan joudutaan turvautumaan ainoastaan likimääräisiin ratkaisuihin.

3 Morozovin diskrepanssiperiaate (discrepancy principle)

Oletetaan, että meillä on ainoastaan jonkinlainen arvio virheen normista:

$$
||\varepsilon|| \leq e,
$$

eli ratkaisulle x pätisi

$$
||Ax - m|| \leq e.
$$

Tällöin voidaan sanoa, että mikä tahansa likimääräinen TSVD-estimaatti $\hat{x}^{(k)}$, jolle pätee

$$||A\hat{x}^{(k)} - m|| \leq e$$

on käytettävissä olevan informaation mielessä "tarpeeksi hyvä"ratkaisu. Toisaalta, jos k valitaan liian suureksi, alkaa likimääräinen tulla mukaan myös kohinaa, joten k pitäisi valita niin pieneksi kuin mahdollista.

Määritelmä 3.1 (Morozovin diskrepanssiperiaate). Valitaan k siten, että

$$||A\hat{x}^{(k)} - m|| \leq e, \qquad ||A\hat{x}^{(k-1)} - m|| > e,$$

toisin sanoen k on pienin luku, jolla *diskrepanssin* $A\hat{x}^{(k)} - m$ normi on pienempi kuin mittausvirheen normi.

Toinen (mutta heuristinen) tapa valita luku k on niin sanottu *L-käyrämenetelmä (L-curve method)*: Nyt jono

$$\delta_k = ||A\hat{x}^{(k)} - m||$$

on monotonisesti laskeva, eli

$$\delta_0 = ||m|| \geq \delta_1 \geq \cdots \geq \delta_r = 0.$$

Toisaalta likimääräisratkaisun $\hat{x}^{(k)}$ normi

$$||\hat{x}^{(k)}||^2 = ||\sum_{i=1}^{k} \frac{1}{\sigma_i}\langle m, u_i\rangle v_i|| = \sum_{i=1}^{k} \frac{1}{\sigma_i^2}|\langle m, u_i\rangle|^2$$

on parametrin k suhteen kasvava jono. Lisäksi kaavasta

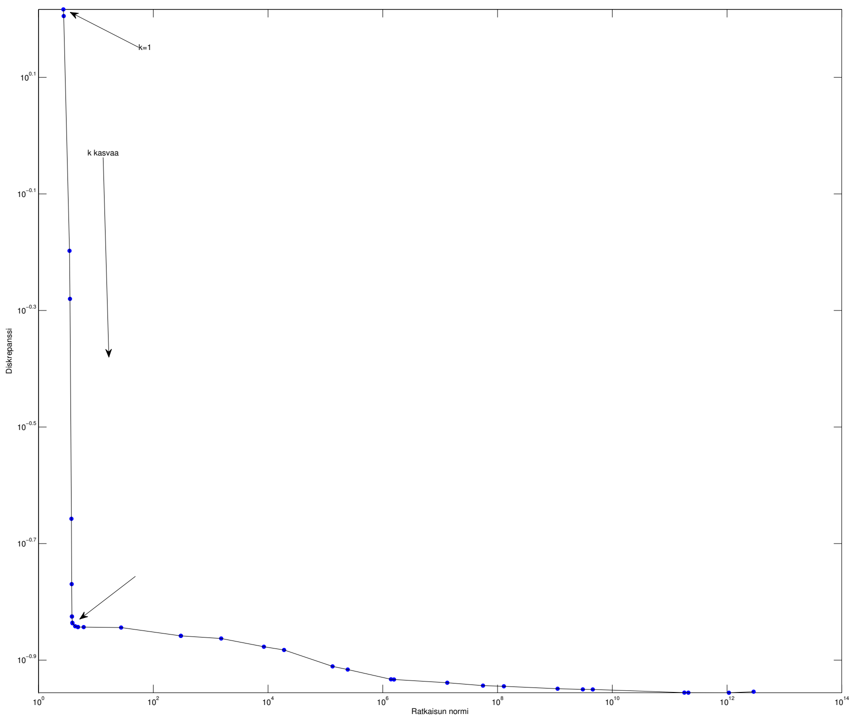
$$||\hat{x}^{(k)}||^2 = \sum_{i=1}^{k} \left(x_{v_i} + \frac{\langle \varepsilon, u_i\rangle}{\sigma_i} \right)^2$$

nähdään, että virhetermi alkaa dominoida, kun $\sigma_i \to 0$. Jos piirretään kuvaaja $k \mapsto (\log ||\hat{x}^{(k)}||, \log ||A\hat{x}^{(k)} - m||)$, saataa[1] se näyttää esimerkiksi sellaiselta kuin kuvassa **??**. Kuvaajasta voidaan nähdä, että L-kirjaimen "kulmassa"

1. diskrepanssi ei enää merkittävästi laske, mutta
2. ratkaisun normissa virhetermi alkaa dominoida.

Toisin sanoen, valitse k L-kirjaimen kulman kohdalta!

[1] Käyrän muoto riippuu ongelmasta, eikä se aina ole L:n muotoinen.



Kuva 1: L-käyrä