Unfolding - uživatelský manuál

Bc. Martin Veselý

Fakulta jaderná a fyzikálně inženýrská

Katedra softwarového inženýrství v ekonomii
Skupina aplikované matematiky a stochastiky při katedře matematiky

Obsah

- motivace z kvantové fyziky
- primitivní unfolding
- unfolding pomocí empirické CDF
- polynomiální regrese
- realizace v MATLABu
- porovnání metod

Motivace z kvantové fyziky

- při měření na subatomární úrovni samotný akt měření ovlivní vlastnosti měřeného objektu
- míra neurčitosti je dána Heisenbergovými relacemi, např. 1. relace (x - poloha, p - hybnost):

$$\Delta p \Delta x \geq \frac{h}{4\pi}$$

- provedeme-li tedy sérii měření, získáme pro stejný objekt několik rozdílných výsledků
- stejné problémy nastanou při měření různých objektů, které jsou však ve stejném kvantovém stavu
- na první pohled se zdá, že máme před sebou několik různých entit
- šum je nutné odstranit ⇒ **UNFOLDING**

Význam unfoldingu pro náhodné matice

- náhodné matice jsou také chaotické systémy
- vznikají jevy podobné jevům v systémech kvantových $\Rightarrow$ neurčitost
- spektra dvou různých matic téhož typu se mohou natolik lišit, že matice budeme považovat za rozdílné
- šum ze spekter je opět nutné odstranit
- **unfolding** převede vlastní čísla na realizace rovnoměrně rozdělené veličiny
- na první pohled paradox, neboť by mělo dojít ke ztrátě veškeré informace (rovnoměrné rozdělení má max. entropii)

Primitivní unfolding

- spektra náhodných matic jsou popsána tzv. Wignerovým polokruhovým zákonem

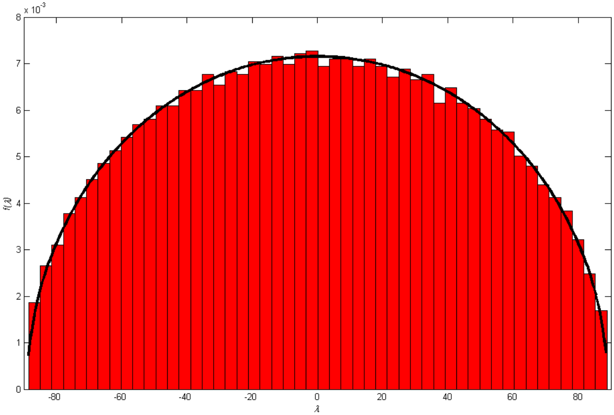
$$f(\lambda) = \theta(\rho(\mathbf{A}) - |\lambda|)\, \frac{2}{\pi[\rho(\mathbf{A})]^2} \sqrt{[\rho(\mathbf{A})]^2 - \lambda^2}$$

- pokud spektrum ořízneme tak, že zachováme tu část kružnice, která je téměř „konstantní“, dostáváme rovnoměrně rozdělenou veličinu
- ořez provedeme zanedbáním vlastních čísel, pro která

$$|\lambda| > k\rho(\mathbf{A}),$$

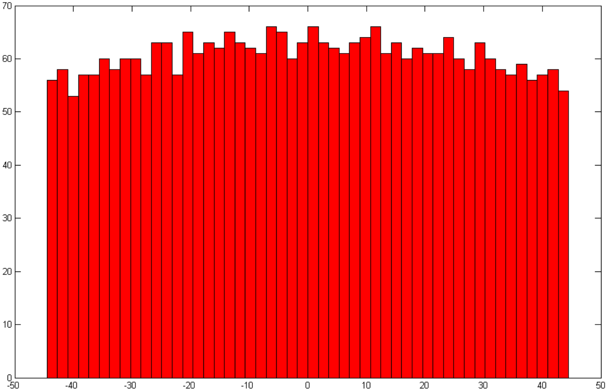
  kde k je většinou $0,5 - 0,75$
- informace o typu matice se zřejmě částečně zachová

Wignerův polokruhový zákon - ukázka



neunfoldované spektrum matice třídy GSE řádu 500

Primitivní unfolding - ukázka



spektrum matice třídy GSE řádu 500 po aplikaci primitivního unfoldingu

Unfolding pomocí empirické CDF - postup

1. mějme dáno spektrum matice, ozn. E
2. určeme $m = \min\{e|e \in E\}$, $M = \max\{e|e \in E\}$ a $n = \text{card}\{e|e \in E\}$
3. interval $\langle m; M\rangle$ rozdělíme na $k = \frac{n}{100}$ intervalů $\langle m_{i-1}; m_i)$, kde $i \in \hat{k}$, $m_0 = m$ a $m_k = M$
4. pro i-tý interval určíme $C_i = C_{i-1} + \text{card}\{e|e \in \langle m_{i-1}; m_i)\}$, $C_1 = \text{card}\{e|e \in \langle m_0; m_1)\}$
5. polynomiální regresí určíme závislost mezi C_i a středem i-tého intervalu
6. získaný regresní polynom aplikujeme na prvky množiny E

Unfolding pomocí empirické CDF - vlastnosti

- metoda „de facto“ hledá regresní funkci, která aproximuje empirickou distribuční funkci spektra (resp. empirickou kumulativní četnost)
- opět přejdeme k rovnoměrnému rozdělení, ale informaci o typu matice v sobě nese distribuční funkce, veškerá nadbytečná informace (šum) je odstraněna
- určíme - li vzdálenosti vlastních čísel unfoldovaného spekra, budou již škálované

Polynomiální regrese - parametry

- regresi provedeme pomocí polynomu $y(x) = \sum_{i=0}^{p} \beta_i x^i$
- odhady parametrů β_i získáme řešením soustavy normálních rovnic

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i & \dots & \sum_{i=1}^{n} x_i^p \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \dots & \sum_{i=1}^{n} x_i^{p+1} \\ \vdots & & & \\ \sum_{i=1}^{n} x_i^p & \sum_{i=1}^{n} x_i^{p+1} & \dots & \sum_{i=1}^{n} x_i^{2p} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \vdots \\ \sum_{i=1}^{n} x_i^p y_i \end{pmatrix}$$

- v našem případě bude $y_i = C_i$ a $x_i = \frac{m_{i-1}+m_i}{2}$ (tj. střed intervalu)
- dále je dobré určit intervaly spolehlivosti (kvůli statistické významnosti parametrů)

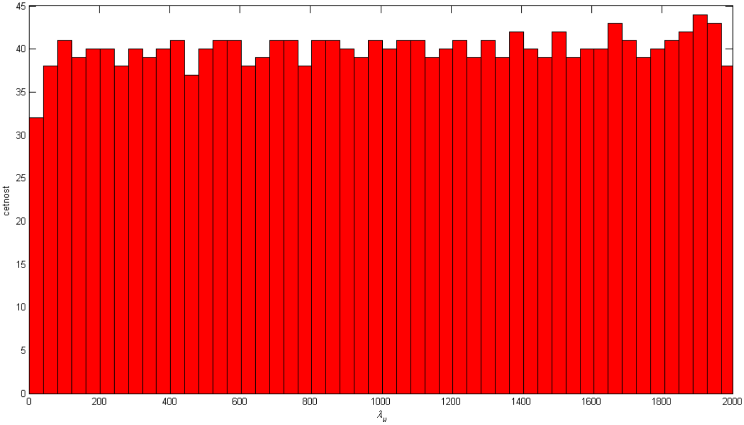
Polynomiální regrese - intervaly spolehlivosti

- $100(1-\alpha)$ % interval spolehlivosti pro parametr b_i má tvar
$$\left(b_i - t_{1-\frac{\alpha}{2}} s_i ; b_i + t_{1-\frac{\alpha}{2}} s_i\right),$$
kde $t_{1-\frac{\alpha}{2}}$ je kvantil Studentova rozdělení s $k-2$ stupni volnosti a s_i je směrodatná odchylka odhadu i-tého parametru
- odhad rozpylu s_i^2 určíme z diagonály kovariační matice parametrů
$$\mathbf{V} = s^2 \left(\mathbf{X}^T \mathbf{X}\right)^{-1},$$
kde matice $\mathbf{X}$ je definována po řádcích $\mathbf{x}_j = (1, x_j, x_j^2 \cdots x_j^p)$ a
$$s^2 = \frac{1}{n-(p+1)} \sum_{\ell=1}^{n} e_\ell^2,$$
kde e_ℓ je ℓ-té reziduum

Realizace v MATLABu

```
function u = unfolding(s)
    N = length(s);
    k = floor(N/100);
    [counting_function, centers] = hist(s,k);
    cumulative_function = zeros(1,length(counting_function));
    cumulative_function(1) = counting_function(1);
    for l = 2:length(counting_function)
        cumulative_function(l) =
           cumulative_function(l-1) + counting_function(l);
    end
    order = 6;
    pol_koef = polyfit(centers,cumulative_function,order);
    u = polyval(pol_koef,s);
```

CDF unfolding - ukázka



spektrum matice třídy GOE řádu 2000 po aplikaci CDF unfoldingu

Porovnání metod - aproximace spacing distribution

- Wignerova domněnka

$$\mathrm{f}_{\mathsf{Wig}}(r) = \theta(r)\mathrm{A}(\beta)r^{\beta}\mathrm{e}^{-\mathrm{B}(\beta)r^2}$$

- Izrailevova formule

$$\mathrm{f}_{\mathsf{Iz}}(r) \approx \theta(r)\mathrm{A}(\beta)\left(\frac{\pi r}{2}\right)^{\beta}\mathrm{e}^{-\frac{\beta\pi^2}{16}r^2-\left(\mathrm{B}(\beta)-\frac{\beta\pi}{4}\right)r}$$

Porovnání metod - výsledky

třída	rozd.	bez unf.	primitivní unf.	CDF unf.
GOE	Wigner	$9,5.10^3$	$9,5.10^3$	$101,7$
	Izrailev	780	$53,8$	$35,4$
GUE	Wigner	$3,9.10^3$	$4,08.10^3$	$47,5$
	Izrailev	$3,8.10^3$	380	$39,2$
GSE	Wigner	$9,6.10^3$	$2,6.10^3$	$35,4$
	Izrailev	$2,7.10^3$	$118,3$	$88,9$

hodnoty Pearsonovy statistiky χ^2 testu pro Wignerovu domněnku a Izrailevovu formuli po aplikaci různých typů unfoldingu (pozn. počet stupňů volnosti: 8)

DĚKUJI ZA POZORNOST