Variante: l_1-Penalisierung anstatt l_1-Nebenbedingung.

[Generelles Paradigma: Ersetze schwierige l_0-Bedingung durch konvexe l_1-Penalisierung.]

(****): für $\rho > 0$ definiere

$$\max \langle \Sigma, Z \rangle - \rho \langle J, |Z| \rangle$$
$$\mathrm{Tr}(Z) = 1$$
$$Z \in S^D_{\geq 0}.$$

(****) ist ein konisches Optimierungsproblem:

$$\max \quad \langle \Sigma, Z \rangle - \rho \sum_{i \leq j} \lambda_{ij} - \rho \sum_{i \leq j} \mu_{ij}$$
$$Z \in S^D_{\geq 0}, \quad (\lambda_{ij}) \in \mathbb{R}^{\binom{D+1}{2}}_{\geq 0}, \quad (\mu_{ij}) \in \mathbb{R}^{\binom{D+1}{2}}_{\geq 0}$$

$$\lambda \quad \langle I, Z \rangle = 1$$

$$u_{ij} \quad \langle Z, E_{ij} \rangle - \lambda_{ij} + \mu_{ij} = 0, \qquad i \leq j.$$

Dualisieren liefert

$$\min \quad \lambda$$

$$\lambda I + \sum u_{ij} E_{ij} - \Sigma \succeq 0$$

$$-u_{ij} + \rho \geq 0 , \quad i \leq j$$

$$u_{ij} + \rho \geq 0 , \quad i \leq j$$

$$= \min \quad \lambda_{max} (\Sigma + U)$$

$$|U_{ij}| \leq \rho , \quad \forall i, j , \quad i \leq j .$$

$\boxed{\text{„robustes" Eigenwertproblem für } \Sigma .}$

<u>Beispiel</u>: von Zhang, et. al. 2012. „Sparse PCA: Convex relaxations, Algorithm and Application".

1288 News articles New York Times

mit 86500 Wörter, d. h.

$$X \in \{0,1\}^{86500 \times 1288}$$

erste dünnbesetzte Hauptkomponente:

$k=2$	$k=3$	$k=9$	$k=14$
united	american	washington	international
states	united	american	world
	state	administration	will
		united	washington
		state	american
		president	administration
		obama	united
		countries	state
		nation	president
			obama
			countries
			nations
			policy
			nuclear.

§3 Das Johnson-Lindenstrauss Lemma

Lemma (Johnson-Lindenstrauss, 1984)
Sei $\varepsilon > 0$, $X \in \mathbb{R}^{N \times N}$, $M \geq C \varepsilon^{-2} \log N$ für eine genügend große Konstante C. Dann gibt es eine Abb. $f : \mathbb{R}^N \to \mathbb{R}^M$, so dass

$$(1-\varepsilon)\,\|x^i - x^j\| \leq \|f(x^i) - f(x^j)\| \leq (1+\varepsilon) \cdot \|x^i - x^j\|$$

für alle $i, j \in \{1, \dots, N\}$ gilt.

Das Lemma wird aus dem folgenden Satz folgen:

Satz Seien $N, M \in \mathbb{N}$ und $\varepsilon \in (0,1)$ gegeben. Definiere die zufällige lineare Abbildung $T : \mathbb{R}^N \to \mathbb{R}^M$ durch die Matrix $T = \frac{1}{\sqrt{M}} (Z_{ij}) \in \mathbb{R}^{M \times N}$, wobei $Z_{ij} \sim N(0,1)$ standardnormalverteilt ist. Dann gilt für jeden Vektor $x \in \mathbb{R}^N$:

$$\Pr\left[(1-\varepsilon)\|x\| \leq \|Tx\| \leq (1+\varepsilon)\|x\|\right] \geq 1 - 2e^{-c\varepsilon^2 M},$$

für eine Konstante $c > 0$.

Bew.: (Lemma)

Wende den Satz auf die $\binom{N}{2}$ Vektoren $x^i - x^j$, $i \leq j$, an. Dann gilt

$$\Pr\left[\exists\, i,j : \neg\left((1-\varepsilon)\|x^i - x^j\| \leq \|Tx^i - Tx^j\| \leq (1+\varepsilon)\|x^i - x^j\|\right)\right]$$

$$\leq \binom{N}{2} 2 e^{-c\varepsilon^2 M} \leq \binom{N}{2} 2 e^{-c\varepsilon^2 C \varepsilon^{-2} \log N}$$

Satz & Ungleichung

$\Pr\left[\bigcup_k A_k\right] \leq \sum_k \Pr[A_k]$

$$= \binom{N}{2} 2 N^{-cC} \leq \frac{1}{2}$$

für geeignete Wahl von C (in Abhängigkeit von c).

Also ist die Wahrscheinlichkeit

$$\Pr\left[\forall\, i,j : (1-\varepsilon)\|x^i - x^j\| \leq \|Tx^i - Tx^j\| \leq (1+\varepsilon)\|x^i - x^j\|\right] \geq \frac{1}{2},$$

insbesondere echt positiv. Also existiert eine lineare Abb. $T = f$ mit der gewünschten Eigenschaft. ☒

Bem.: Die Beweisstrategie, die Existenz von mathematischen Objekten mit Hilfe von "$\Pr[\ldots] > 0$" zu zeigen, geht auf Erdős zurück. Sie heißt die probabilistische Methode. (ungemein wirkungsvoll!)

Erinnerung: Normalverteilung

$X \sim N(0,1)$ Standardnormalverteilte Zufallsvariable

Dichtefkt. $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ auf $\mathbb{R}$.

<u>allgemein:</u> $X \sim N(\mu, \sigma^2)$ normalverteilt mit

Erwartungswert μ, Varianz σ^2,

Dichtefkt. $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2/2}$.

D.h. $E[X] = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2/2} dx = \mu$

$$\mathrm{Var}[X] = E\left[(X - E[X])^2\right] = \sigma^2.$$

<u>wichtige Eigenschaft:</u> $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_1, \sigma_1^2)$ unabhängig, $\alpha_1, \alpha_2 \geq 0$. Dann gilt

$$\alpha_1 X_1 + \alpha_2 X_2 \sim N\left(\alpha_1 \mu_1 + \alpha_2 \mu_2,\ (\alpha_1 \sigma_1)^2 + (\alpha_2 \sigma_2)^2\right).$$

<u>Lemma</u> (Markov Ungleichung)

Sei X eine Zufallsvariable in $\mathbb{R}$ mit Wahrscheinlichkeitsmaß μ. Sei $h : \mathbb{R} \to [0, \infty)$ monoton wachsend. Dann gilt für $a \in \mathbb{R}$: $h(a) \Pr[X \geq a] \leq E[h(X)]$.