Zunächst zum Beweis des Theorems von Fan.

Wdh. Theorem (Fan):

Sei $C \in S^n$ sym. Matrix mit Eigenwerten $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$

Dann gilt: $\lambda_1 + \ldots + \lambda_k = \max\{\operatorname{Tr}(CBB^T) : B \in \mathbb{R}^{n\times k}, B^TB = I_k\}$

$$\overset{(*)}{=} \max\{\operatorname{Tr}(CY) : Y \in S^n, \operatorname{Tr}(Y) = k, Y \succeq 0, I - Y \succeq 0\}.$$

Wdh.: → OR (2015)

Satz (Minkowski, Krein-Milman)

Sei $D \subseteq \mathbb{R}^n$ eine kompakte und konvexe Menge. Dann gilt

$D = \operatorname{conv}(\operatorname{ext}(D))$, wobei $\operatorname{ext}(D) = \{x \in D : x$ sind Extrempunkte von $D\}$.

Lemma Definiere

$$K_1 := \{Y \in S^n : \operatorname{Tr}(Y) = k, I_n \succeq Y \succeq 0\}$$

$$K_2 := \{BB^T : B \in \mathbb{R}^{n\times k}, B^TB = I_k\}$$

Dann gilt $\operatorname{ext}(K_1) = K_2$, insbesondere $\operatorname{conv}(K_2) = K_1$.

Beweisskizze:

Sei $P \in O(n)$ dann gilt

$$Y \in \operatorname{ext}(K_1) \Leftrightarrow PYP^T \in \operatorname{ext}(K_1).$$

Definiere nun das Polytop

$$Q := \{ x \in [0,1]^n : e^T x = k \}, \text{ wobei } e = (1,\dots,1)^T.$$

Mittels der Spektralzerlegung von $Y \in S^n$ folgt

$$Y = PDP^T \quad \text{mit } P \in O(n) \text{ und Diagonalmatrix } D := \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}.$$

Dann gilt $Y \in ext(K_1) \Leftrightarrow D \in ext(K_1) \Leftrightarrow d \in ext(Q)$.

Außerdem, $ext(Q) = \{ x \in \{0,1\}^n : e^T x = k \}$ und damit

$$|\{ d_i : d_i = 1 \}| = k \ , \ |\{ d_i : d_i = 0 \}| = n-k.$$

Da die d_i die Eigenwerte von Y beschreiben gilt nun

$Y \in ext(K_1) \Leftrightarrow$ Y hat genau k Eigenwerte mit Wert 1 und $n-k$ Eigenwerte mit Wert 0.

Das bedeutet, dass Y eine Projektion auf einen k-dimensionalen Unterraum ist: $\exists\, B \in \mathbb{R}^{n \times k} : B^T B = I_k$ und $Y = BB^T$.

$\square$

Dies nutzen wir nun um Fans Theorem zu beweisen.

Beweisskizze (Fan):

Aus OR folgt (*) $\hat{=} \max\{c^T x : x \in D\} = \max\{c^T x : x \in \operatorname{ext}(D)\}$.

Sei p^* dieses Maximum.

$\lambda_1 + \dots + \lambda_k \leq p^*$: Sei $C = \sum_{i=1}^{n} \lambda_i \mu_i \mu_i^T$, dann ist $B = [\mu_1 \dots \mu_k] \in \mathbb{R}^{n \times k}$ und $BB^T \in K_2$, damit zulässig für das zugehörige SDP mit Zielfunktionswert $\operatorname{Tr}(CBB^T) = \lambda_1 + \dots + \lambda_k$.

$\lambda_1 + \dots + \lambda_k \geq p^*$: Nutze erneut Spektralzerlegung von C:

$C = PDP^T$, $P \in O(n)$, D Diagonalmatrix, dann gilt für $BB^T \in K_2$:

$$\operatorname{Tr}(CBB^T) = \dots = \operatorname{Tr}(D \underbrace{P^T B (P^T B)^T}_{\in K_2})$$

Das bedeutet:

$$p^* = \max\{\operatorname{Tr}(DM) = \sum_{i=1}^{n} \lambda_i M_{ii} : M \in K_2\}$$

Für $M \in K_2$ liegt $\operatorname{diag}(M) = (M_{11}, \dots, M_{nn})$ in Q.

Daher gilt $p^* \leq \max\{\sum_{i=1}^{n} \lambda_i x_i : x \in Q\} = \lambda_1 + \dots + \lambda_k$ □

Aus letzter Vorlesung wissen wir, dass man mit Hilfe der Eigenwerte $\lambda_1, \dots, \lambda_D$ und der Eigenvektoren $b_1, \dots, b_D$ von skalierten Varianten der Kovarianzmatrix $\underbrace{}_{(\hat{=} XX^T)}$ Informationen über die Struktur der zugrundeliegenden Daten erhält.
Dass D gegebenenfalls sehr groß sein kann wird an folgendem Beispiel deutlich.

<u>Bsp.</u> Man betrachte einen Datensatz von $N = 500$ Bildern der Größe 1000×1000 Pixel.
Diese kann man auffassen als Vektoren $x_1, \dots, x_N$ der Größe $1000 \cdot 1000 = 10^6 = D$.
Die zugehörige (skalierte-) Kovarianzmatrix $XX^T \in \mathbb{R}^{D \times D}$ müsste nun in Eigenwerte / Eigenvektoren zerlegt werden, wofür moderne Algorithmen eine Laufzeit der Größenordnung $O(D^3) = O(10^{18})$ benötigen. $\rightarrow$ Ansatz: Dimensionsreduktion

Wissen: $\operatorname{rang}(XX^T) \leq N$, hier: $\operatorname{rang}(XX^T) = N$, d.h. $\lambda_1, \dots, \lambda_N > 0$, (andernfalls betr. l.u. Teilmenge v. x_i)

$$XX^T \cdot (b_1, \dots, b_N) = \lambda_1 b_1 + \dots + \lambda_N b_N = B \cdot \Lambda \Rightarrow X^T X X^T B = X^T B \cdot \Lambda \quad (*)$$

Definiere $\tilde{B} = X^T B$, dann ist $(*)$ äquivalent zu

$$X^T X (\tilde{b}_1, \dots, \tilde{b}_N) = X^T X \tilde{B} = \tilde{B}\Lambda = \lambda_1 \tilde{b}_1 + \dots + \lambda_N \tilde{b}_N,$$ d.h. man kann

mit Aufwand $O(N^3)$ die Eigenwerte λ_i und Eigenvektoren $\tilde{b}_i \in \mathbb{R}^N$ der Matrix $X^T X \in \mathbb{R}^{N \times N}$ berechnen.

Weiterhin löst $B = X\tilde{B}\Lambda^{-1}$ die Eigenwertproblemgleichung zu XX^T, da

$$XX^T B = XX^T X \tilde{B} \Lambda^{-1} = X \tilde{B} \Lambda \cdot \Lambda^{-1} = X\tilde{B} = B\Lambda .$$

Damit ergibt sich, dass man die Eigenwerte λ_i mit Aufwand $O(N^3)$ und die zugehörigen Eigenvektoren b_i mit Aufwand $O(DN^2)$ berechnen kann. (Die Berechnung von $\underbrace{X}_{\in \mathbb{R}^{D\times N}} \underbrace{\tilde{B}}_{\in \mathbb{R}^{N\times N}} \underbrace{\Lambda^{-1}}_{\in \mathbb{R}^{N\times N}} \in O(DN^2)$).

Da $N \ll D$ $(500 \ll 10^6)$ in unserem Beispiel ergibt sich eine Laufzeit, welche proportional zu $500^2 \cdot 10^6 = 25 \cdot 10^{11}$ ist anstatt proportional zu 10^{18}.