

# Lecture 13 - scribbles - polytopes

Friday, February 17, 2017  
15:20

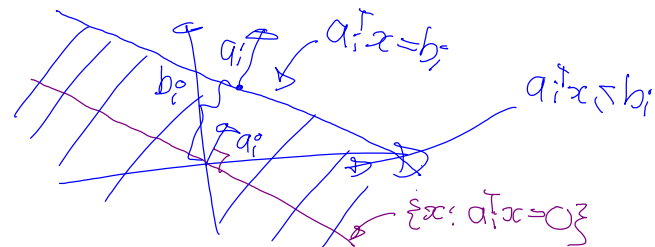
today:

- convex polytopes
- BCFW

Convex polytope:

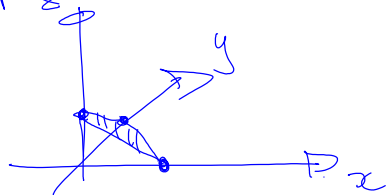
polytope  $\triangleq$   $\left\{ \begin{array}{l} \text{convex hull of finite \# of points} \quad (V\text{-representation}) \\ \text{bounded polyhedron} \end{array} \right.$   
 $\rightarrow$  intersection of finite  $\#$  of half-spaces  $(H\text{-representation})$   
 $\{x: Ax \leq b\}$

$A$  is  $m \times d$   
 row of  $A$  e.g.  $a_i^T x \leq b_i$  suppose  $\|a_i\|_2 = 1$   
 in dimension  $d$ ,  $\{a_i^T x = b_i\} \rightarrow d-1$  dim object



intersection of  $r$  hyperplanes gives  
 you a  $d-r$  dimensional  
 object

probability simplex in dimension  $d = \text{convex hull}(\{e_i\}_{i=1}^d)$



$$\left[ \begin{array}{l} \sum_{i=1}^d x_i = 1, \quad x_i \geq 0 \end{array} \right] H\text{-representation}$$

suppose polytope is  $d$ -dimensional

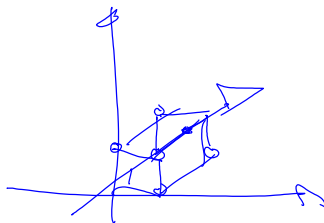
whole polytope  $\rightarrow d$ -dim face  
facet  $\rightarrow d-1$  dimensional face  
 ridge  $\rightarrow d-2$  " " "

face of a polytope = intersection of polytope with  
 several hyperplanes which  
 determine its boundary

ridge  $\rightarrow d-2$  " " "  
 edge  $\rightarrow 1$  dimensional face  
 vertex  $\rightarrow 0$  " face

example: • unit (hyper)-cube

(scaled translated  
 version of  $\ell_\infty$ -ball  
 $\|x\|_\infty \leq 1$ )

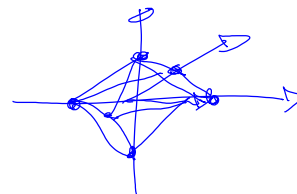


$2^d$  vertices

convex-hull  $(\{0,1\}^d)$

2d facets:  $\{x_i \geq 0, i=1, \dots, d\}$   
 $x_i \leq 1$

•  $\ell_1$ -ball  $\{x: \|x\|_1 \leq 1\}$   $2^d$  facets  
 but  $2d$  vertices  
 $\text{conv}(\{\pm e_i\}_{i=1}^d)$



• simplex  $\Delta_d$ :  $d$  vertices  $\rightarrow e_i$   
 $d$  facets  $\rightarrow x_i = 0$

• in structured SVM, we have  $\Delta_{i=1}^d (r_i)$  (product structure  $\Leftrightarrow$  "cube like")

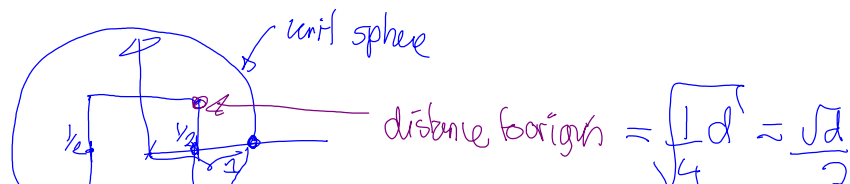
in simplex:  $\|x\|_1 = 1$  always  
 but  $\|x\|_2$  as small as

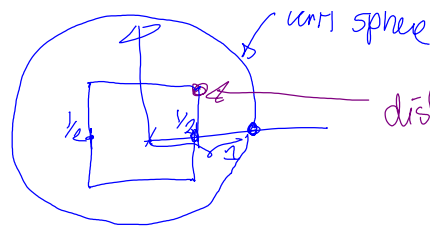
$$\left\| \frac{1}{d} \mathbf{1} \right\|_1 = 1, \quad \left\| \frac{1}{d} \mathbf{1} \right\|_2 = \sqrt{\underbrace{\frac{1}{d} + \dots + \frac{1}{d}}_d} = \frac{1}{\sqrt{d}}$$

$\|x\|_1 \geq \|x\|_2 \geq \|x\|_\infty$  always

( $\|\cdot\|_p \geq \|\cdot\|_q$  if  $p < q$ )

\* pitfalls of low-dimensional intuition in high dimension

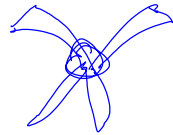




$$\text{distance for origin} = \sqrt{\frac{1}{4}d} = \frac{\sqrt{d}}{2}$$

$$d=4 \rightarrow \frac{2}{2} = 1$$

in high dimension -



volume of sphere of radius  $r$  in dim  $d$

$$= \frac{\pi^{d/2}}{\Gamma(\frac{d}{2} + 1)} r^d$$

• volume of unit sphere  $\rightarrow 0$  as  $d \rightarrow \infty$

• whereas cent-where volume = 1

Affine invariance of FW:

recall (lecture 10)  $\tilde{M} \xrightarrow{\tilde{A}} M$

$$\tilde{f}(\tilde{x}) = f(\tilde{A}\tilde{x})$$

$\tilde{x} \in M$

running FW on  $\tilde{M} \ni \tilde{f}$

equivalent to running on  $M \ni f$

$$x_t = \tilde{A}\tilde{x}_t$$

well let  $\tilde{M}_i = \Delta_{|S_i|}$  let  $\tilde{A}$  be marginalization operation

$$\mu_c(y_c) = \sum_{y \vee y_c} \alpha(y) = a_{c,y_c}^T \alpha$$

marginal polytype for example

$$\mu_c = \tilde{A}\tilde{\mu}_c$$

Marginal polytype:

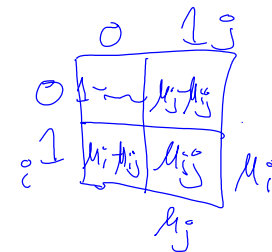
\* need to decide what "parameterization" of marginals you want to use ...

e.g. 1) simplest: node & edge marginals # of possible values  $\rightarrow \infty$  (redundant)

$\mu_i(y_i) \rightarrow |V| \cdot k$  variables (no need of have triangulated graph)  
 $\mu_{ij}(y_i, y_j) \rightarrow |E| \cdot k^2$  variables

2) "minimal" representation  $\rightarrow |V|(k-1) + |E|(k-1)^2$

for example for  $k=2$  (binary variable)  $\mu_i, \mu_j, \mu_{ij}$



3) use  $\mu_c(y_c)$  with maximal cliques

getting ft representation of Marginal polytope :

Fourier-Motzkin elimination

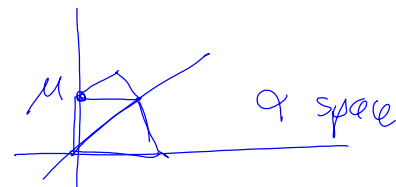
start with  $\Delta_{|Y|} \rightarrow \sum_y \alpha(y) = 1$  and  $\alpha(y) \geq 0 \forall y \in Y$   
 $B\alpha \leq c$

if has  $a_i^T x = b_i$   $\begin{cases} a_i^T x \geq b_i \\ a_i^T x \leq b_i \end{cases}$

add constraint:  $\tilde{A}\alpha = \mu$   
 $\rightarrow \tilde{A}\alpha - \mu = 0$

to get  $M$ , the marginal polytope

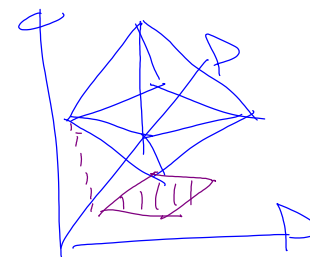
you want to eliminate  $\alpha$  variable



from the description

$$\begin{pmatrix} B & 0 \\ \tilde{A} & -I \end{pmatrix} \begin{pmatrix} \alpha \\ \mu \end{pmatrix} \begin{matrix} \leq c \\ = 0 \end{matrix}$$

first, make  $\tilde{A}\alpha = \mu$  as inequalities  $\tilde{A}\alpha \leq \mu$  &  $\tilde{A}\alpha \geq \mu$



Fourier-Motzkin elimination : can we want to eliminate  $\alpha_1$

feas if  $a_{1d} > 0$

any you want to minimize is

$$\sum_i a_{ij} x_j \leq b_i \Rightarrow x_d \leq \frac{b_i - \sum_{j \neq d} a_{ij} x_j}{a_{id}}$$

$$\begin{matrix} \dots \\ \dots \\ \dots \\ n_a \end{matrix} \leq x_d \leq \begin{matrix} \dots \\ \dots \\ \dots \\ n_b \end{matrix} \xrightarrow{\text{inequalities}} n_a \cdot n_b \text{ new inequalities} \rightarrow \text{quadratic explosion}$$

say start with  $\frac{n}{2}, \frac{n}{2}$  go from  $n$  inequalities to  $\frac{n^2}{4}$   
after  $d$  steps ...  $\frac{n^2 d}{4}$

back to FW for SVM struct:

what about  $C_f$ ?

$$C_f = \sup_{\substack{\gamma \in [0,1] \\ s, x \in M \\ x_\gamma = x + \gamma(s-x)}} \frac{1}{2} (f(x_\gamma) - f(x) - \langle Df(x), x_\gamma - x \rangle)$$

$$\text{by Taylor } f(x_\gamma) = f(x) + \langle Df(x), x_\gamma - x \rangle + \underbrace{\frac{1}{2} (x_\gamma - x)^T H(\tilde{x}(\gamma)) (x_\gamma - x)}_{\text{remainder}} \\ \tilde{x}(\gamma) \in [x, x_\gamma] \\ x_\gamma - x = \gamma(s-x)$$

$$C_f \leq \sup_{s, x \in M} (s-x)^T H(s-x) \quad \text{for SVM struct}$$

$$\text{Hessian } H = \lambda A^T A, \quad M = \sum_{i=1}^n \Delta p_i$$

$$\Rightarrow C_f \leq \lambda \sup (a-b)^T A^T A (a-b)$$

$$\alpha, \beta \in M \quad \frac{1}{\|A(\alpha - \beta)\|_2^2}$$

$$\begin{aligned} \|A(\alpha - \beta)\| &= \left\| \frac{1}{n} \sum_{i=1}^n \psi_i(y) (\alpha_i(y) - \beta_i(y)) \right\| \\ &\leq \frac{1}{n} \sum_{i=1}^n \underbrace{\|\psi_i(y)\|}_{\leq R} (\alpha_i(y) + \beta_i(y)) \\ &\leq \frac{1}{n} R n(2) = 2R \end{aligned}$$

$$C_f \leq \frac{4R^2}{\lambda}$$

let's compare to  $C_f \leq L_{\|\cdot\|} \text{diam}_{\|\cdot\|}^2$  bound; let's use  $\|\cdot\|_2$   $\text{diam}(M)^2 = 2n$

Lipschitz constant in  $\ell_2$ -norm = largest e-value of Hessian

$$\lambda A^T A = \int_{\mathcal{Y}} \sum_{i,j} \langle \psi_i(y), \psi_j(y) \rangle \mathbf{e}_i \mathbf{e}_j^T$$

say e.g.  $\langle \psi_i(y), \psi_j(y) \rangle \approx 1$  for lots of outputs

$$\mathbb{E}[\mathbf{I}] = \mathbf{I} \cdot d$$

here, get huge bound for useless

exercise to the reader: if you use  $\|\cdot\|_{\ell_{\infty,1}}$  block-norm on  $M$

$$\text{i.e. } \|\alpha\|_{\ell_{\infty,1}} = \max_{i=1,\dots,n} \|\alpha_i\|_1$$

$$\text{diam}_{\|\cdot\|_{\ell_{\infty,1}}}(M)^2 = 2$$

$$L_{\|\cdot\|_{\ell_{\infty,1}}} = \frac{2R^2}{\lambda}$$

$L$  was such that

$$\|Df(x) - Df(y)\|_* \leq L \|x - y\|$$

$$(\|\cdot\|_{p,q})^* = \|\cdot\|_{q,p}$$

**Pointers:**

- to learn more about polytopes:
    - book: Lectures on Polytopes, Ziegler, 1995
  - to learn more about the marginal polytope:
    - see Section 3.4 of:
 

M. J. Wainwright and M. I. Jordan, Graphical models, exponential families, and variational inference. Foundations and Trends in Machine Learning

[http://www.eecs.berkeley.edu/~wainwrig/Papers/WaiJor08\\_FTML.pdf](http://www.eecs.berkeley.edu/~wainwrig/Papers/WaiJor08_FTML.pdf)
- > also, see p.80 Figure 4.1 for an example of fractional corners for the local consistency polytope (vs. the marginal polytope which only has integer vertices)