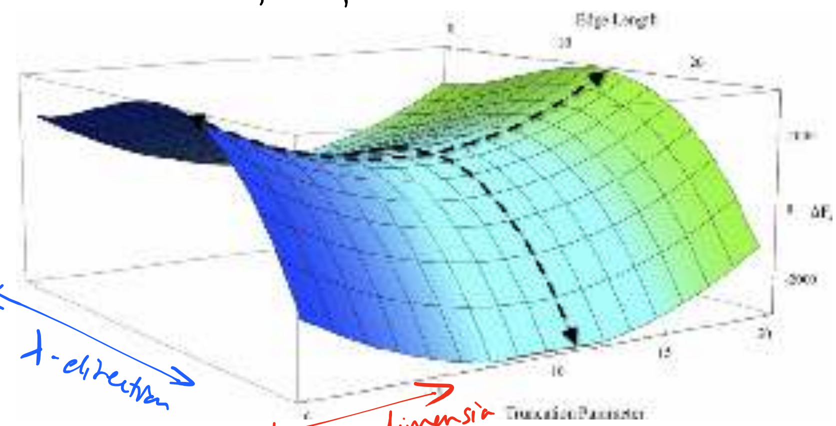
Saddle Point Interpretation, shared Lagrangians

Monday, March 1, 2021 10:09 AM

Ex: $f(x, \lambda) = x^2 - \lambda^2$



λ-direction

x dimension

Saddle-points:

sometimes we want them (primal-dual optim.)

sometimes we don't (e.g., stationary pt. w/ indefinite Hessian)

§4.2 in B+V

we don't mean this 1D interpretation

(This is an unwanted saddle)

Primal

$$p^* = \min_{\substack{x \\ x\in D}} f_0(x) \quad f_i(x) \le 0,\ Ax=b \quad = \min_{x\in D} f_0(x) + \underbrace{\sup_{\substack{\lambda \ge 0 \\ \nu}} \left(\textstyle\sum_i \lambda_i f_i(x) + \nu^T(Ax-b)\right)}$$

if $f_i(x) > 0$ for some i this is $+\infty$

if $a_i^T x - b_i \ne 0$ for some i this is also $+\infty$

$$= \min_{x\in D} \sup_{\substack{\lambda\ge 0\\ \nu}} \mathcal{L}(x,\lambda,\nu)$$

BTW, another related characterization

if $f(x) = f_0(x) + I_{\{f_i(x)\le 0,\ Ax=b\}}$ is in $\Gamma_0(\mathbb{R}^n)$

by Fenchel-Moreau Thm, $f = f^{**}$

so $p^* = \min_x f(x)$

$= \min_x \max_y \langle x, y\rangle - f^*(y)$

Dual

$$d^* = \max_{\substack{\lambda\ge 0\\ \nu}} g(\lambda,\nu) = \max_{\substack{\lambda\ge 0\\ \nu}} \min_{x\in D} \mathcal{L}(x,\lambda,\nu)$$

Weak-duality: $d^* \le p^*$

$\Leftrightarrow$ "min-max inequality".

ALWAYS TRUE

$$d^* = \max_{\substack{\lambda\ge 0\\ \nu}} \min_{x\in D} \mathcal{L}(x,\lambda,\nu) \le \min_{x\in D} \max_{\substack{\lambda\ge 0\\ \nu}} \mathcal{L}(x,\lambda,\nu) = p^*$$

= if strong-duality holds "strong max-min"

*Note: all "min" should be "inf"
all "max" should be "sup"

Saddle point

... this is when ① strong-duality / strong max-min

(2) inf/sup are achieved

i.e., $(x^*, (\lambda^*, \nu^*))$ is a "saddle-point" of $\mathcal{L}(x, (\lambda, \nu))$

if

$$\left.\begin{aligned} \mathcal{L}(x^*, (\lambda^*, \nu^*)) &= \boxed{\inf_x \mathcal{L}(x, (\lambda^*, \nu^*))} \\ \mathcal{L}(x^*, (\lambda^*, \nu^*)) &= \sup_{\lambda, \nu} \mathcal{L}(x^*, (\lambda, \nu)) \end{aligned}\right] \text{Eq'n}$$

$\Rightarrow$ strong max-min

★ **Implication**: If we know λ^*, ν^* then can find x^*

by solving the unconstrained problem $\min_{x \in D} \mathcal{L}(x, \lambda^*, \nu^*)$

(instead of solving $\min_{\substack{x \in D \\ f_i(x) \le 0 \\ h_i(x) = 0}} f_0(x)$)

Problems w/ Shared Lagrangians

$$\begin{aligned} &\min \ \|x\|_1 \\ &\text{s.t.} \ \underbrace{\|Ax-b\|_2 \le \varepsilon}_{\text{hard to project onto}} \end{aligned} \iff \begin{aligned} &\min \ \|x\|_1 \\ &\underbrace{\|Ax-b\|_2^2 - \varepsilon^2 \le 0} \end{aligned}$$

$$\mathcal{L}(x, \lambda) = \|x\|_1 + \left(\|Ax-b\|_2^2 - \varepsilon^2\right) \cdot \lambda$$

w/ correct λ^*,

equivalent to $\underbrace{\min_x \|x\|_1 + \lambda^* \underbrace{\|Ax-b\|_2^2}_{\text{differentiable}}}_{\text{nicer!}}$

Even if we don't know λ^*...

1) guess λ, solve $x = x(\lambda)$, check constraint

$\underbrace{\text{update } \lambda}_{\text{vague}}$ (i.e., solve dual problem!)

2) often ε is not known (hyper-parameter)

and set via cross-validation

... so do cross-validation on λ directly

This assumes $\exists$ of saddle-pts

(stronger than just strong duality)

Prop: $\left.\begin{aligned} &\text{Slater's on Primal} \\ &\text{and Slater's on Dual} \end{aligned}\right\} \Rightarrow \exists$ of saddle pts

we'll talk about
KKT conditions for
these later