Paper notes

$S = \{s_i : i = 1, \dots\}$

$\text{conv}(S) \supseteq S$

- check if it is zig-zagging
- print gap too
- improve opt Δ
- simplify entropy regularization / norm constraint
- check $\gamma \times$ other things on 1D & 2D examples.
- check distance metrics for prob. dist using Optimal Transport
- Estimate $L / \text{diam}(D)$ etc. on function spaces for known f
- check 1D non Δ behavior of line search, fully corr. & fixed (2D non too)

it was mentioned that it stops at 1^{st} itr.

- see works of Jaggi Julien.
- in FW we select $S = \{s_0, s_1, \dots\}$ but $g_t \in \text{conv}(S)$ do something so that $\text{conv}(D) \supset D$ & is more powerful.

$g_t^{FW} := \langle \nabla f(x^t), x^{(t)} - s_t \rangle$

$= \ell(x^{(t)}) - \ell(s_t)$

$= -\langle \nabla f(x^t), s_t - x^{(t)} \rangle$

$-\langle \nabla f(x^t), x^t - v^t \rangle = \ell(v^t) - \ell(x^t)$

$v^t \in \operatorname{argmax}_{s \in S} \langle \nabla f(x^t), v \rangle$ || ~~most far from grad.~~ (closest to grad.

$x^t - v^t \in \|$ = away from v^t

$s^t - x^t \in \|$ = close to s^t

- DCC, convex or strongly convex?
- Regularization from boosting VI for h
- Approx ~~not~~ diff distribution than θ_{mix} gaussian ⊥ exponential?