SMOOTHING (Nesterov, 2005)

Used for nonsmooth convex problems: $\min f(x)$

approximate by smooth convex problem, $\min f_\mu(x)$ parameter

and use ~~sub~~ gradient descent.

<u>Complexity</u> # iterations depends on $(L_\mu / \varepsilon_\mu)$ linear in this, with careful design of algo.

where L_μ is Lipschitz constant of ∇f_μ

and ε_μ is accuracy with which smooth problem is solved.

<u>Trade-off</u>:

Large $L_\mu \Leftrightarrow$ less smoothing $\Rightarrow$ more accurate approximation

Small $L_\mu \Leftrightarrow$ more smoothing $\Rightarrow$ faster convergence.

<u>Eg</u>: $f(X) = \lambda_{max}(X)$. (Minimize max eval, subject to some constraints)

$X \succeq 0$.

Smooth approximation:

$$f_\mu(X) = \mu \log\left(\sum_{i=1}^{n} e^{\lambda_i(X)/\mu}\right) - \mu \log n.$$

$$\lim_{\mu \to 0} f_\mu(X) = f(X). \qquad L_\mu = \frac{1}{\mu}.$$