INTERIOR POINT METHODS

WHAT'S AN INTERIOR POINT METHOD?

$$\text{minimize} \quad f(x)$$

solve: invert Hessian $\qquad \nabla^2 f(x)\Delta x = -\nabla f(x)$

line search $\qquad x \leftarrow x + \alpha\Delta x$

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \end{aligned}$$

Lagrangian $\qquad L(x, \lambda) = f(x) + \langle \lambda, Ax - b \rangle$

$$\begin{pmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x) \\ b \end{pmatrix}$$

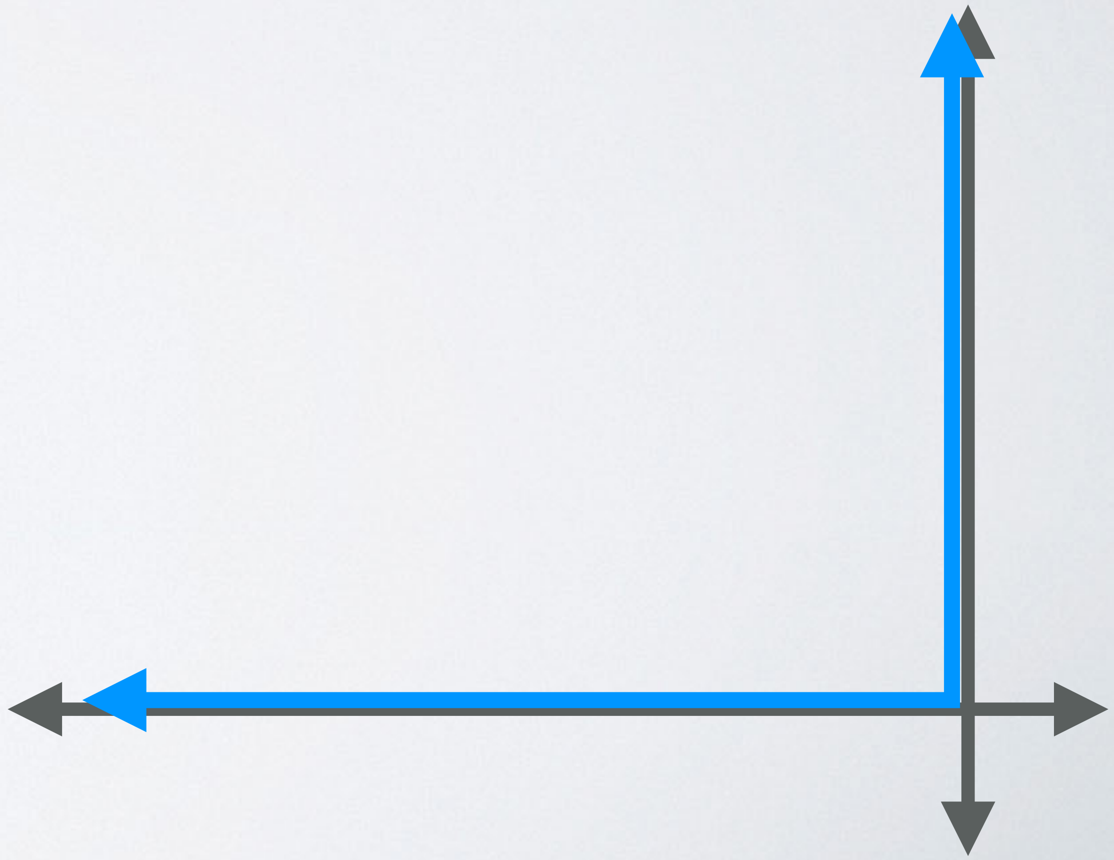
What about inequality constraints?

IMPLICIT FORM CONSTRAINTS

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \\ & g(x) \leq 0 \end{aligned}$$

$$\mathcal{X}_{-}(z) = \begin{cases} 0, & \text{if } z \leq 0 \\ \infty, & \text{otherwise} \end{cases}$$

$$\begin{aligned} \text{minimize} \quad & f(x) + \mathcal{X}_{-}(g(x)) \\ \text{subject to} \quad & Ax = b \end{aligned}$$



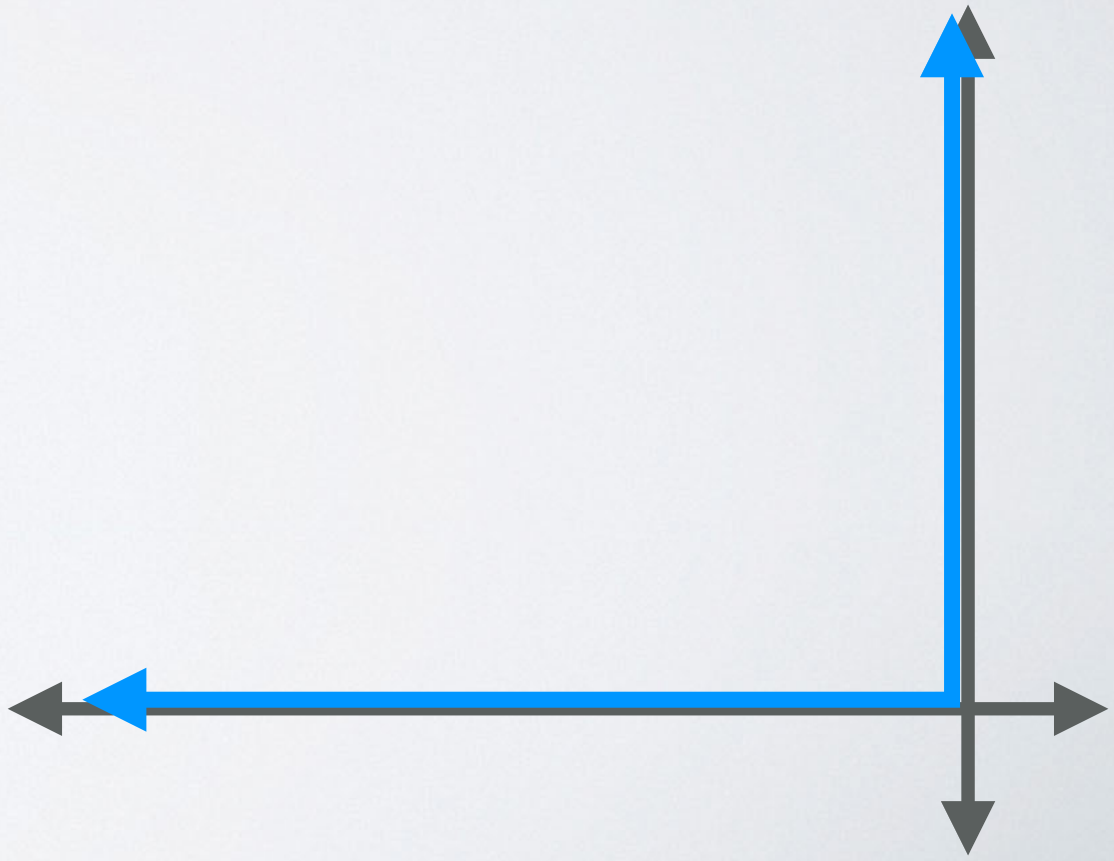
INTERIOR POINT METHODS

$$\begin{array}{ll}\text{minimize} & f(x) + \mathcal{X}_-(g(x)) \\ \text{subject to} & Ax = b\end{array}$$

$$\mathcal{X}_-(z) = \begin{cases} 0, & \text{if } z \leq 0 \\ \infty, & \text{otherwise} \end{cases}$$

IP methods solve this

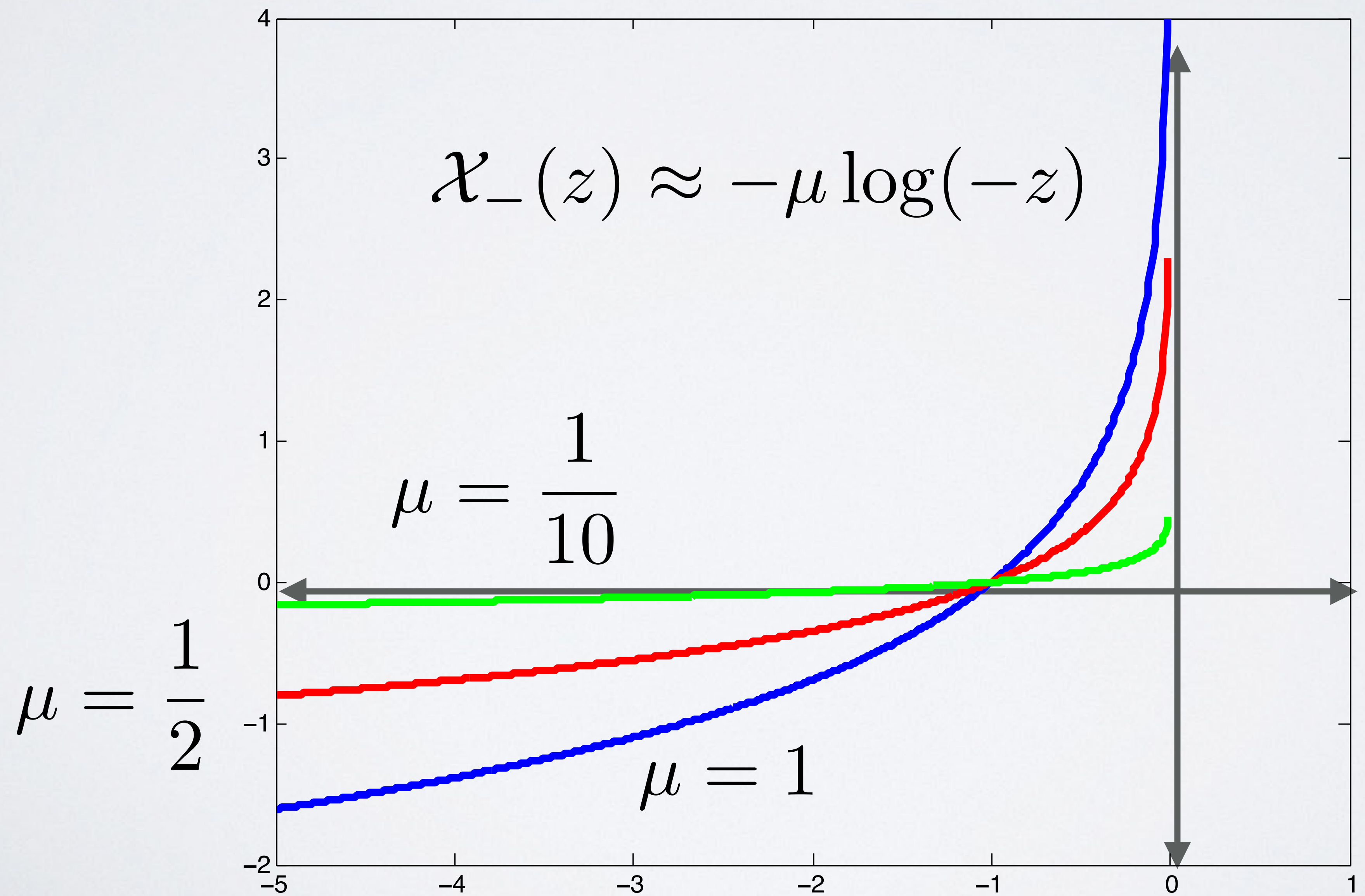
When objective is smooth, we can use constrained Newton, BFGS, CG, etc…



LOG BARRIER FUNCTION

$$\begin{array}{ll} \text{minimize} & f(x) + \mathcal{X}_{-}(g(x)) \\ \text{subject to} & Ax = b \end{array}$$

$$\mathcal{X}_{-}(z) = \begin{cases} 0, & \text{if } z \leq 0 \\ \infty, & \text{otherwise} \end{cases}$$

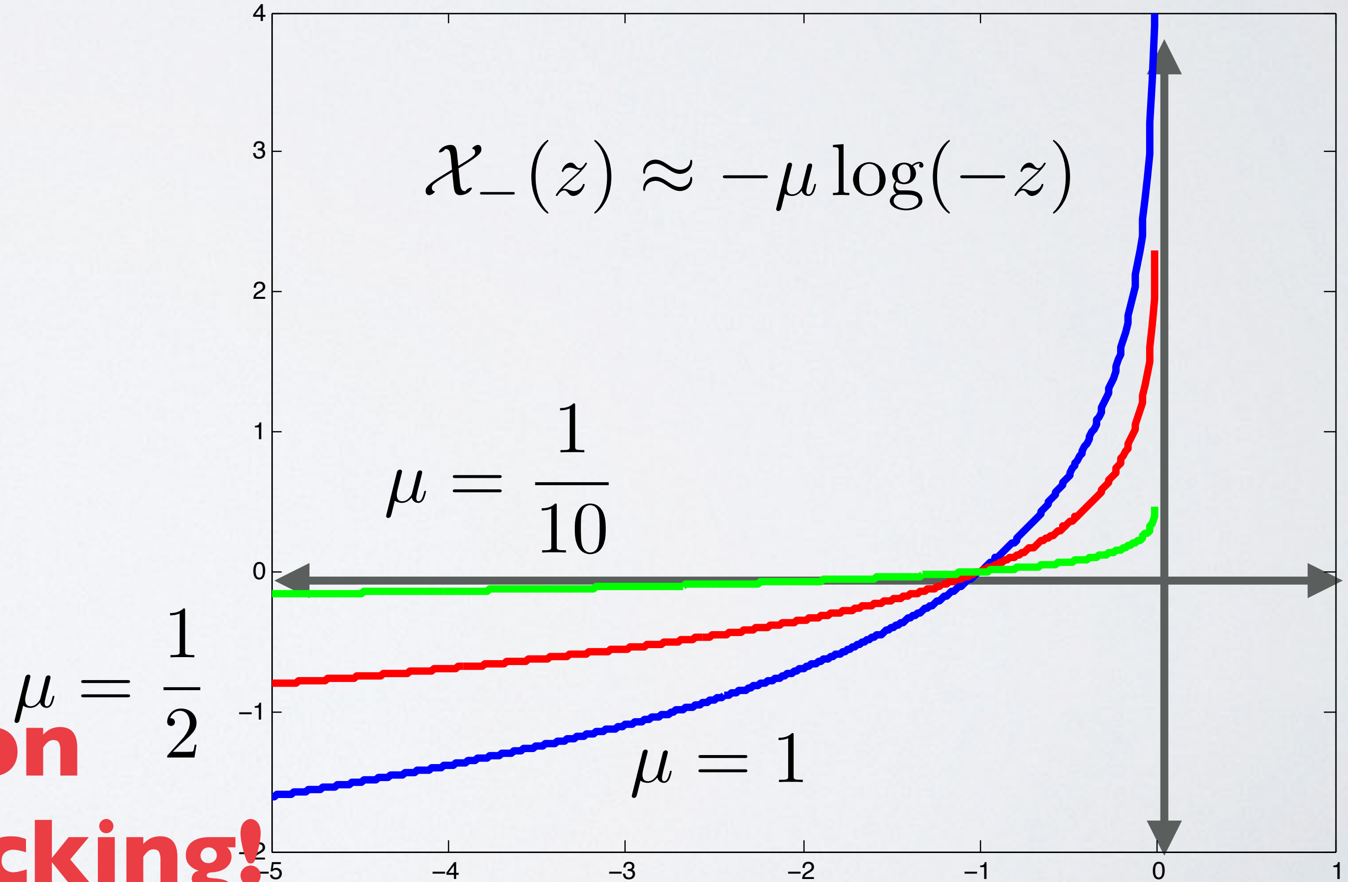


BARRIER METHOD

$$\begin{aligned} \text{minimize} \quad & f(x) - \mu \log(-g(x)) \\ \text{subject to} \quad & Ax = b \end{aligned}$$

Pick some "small" μ

Solve the problem with Newton's method

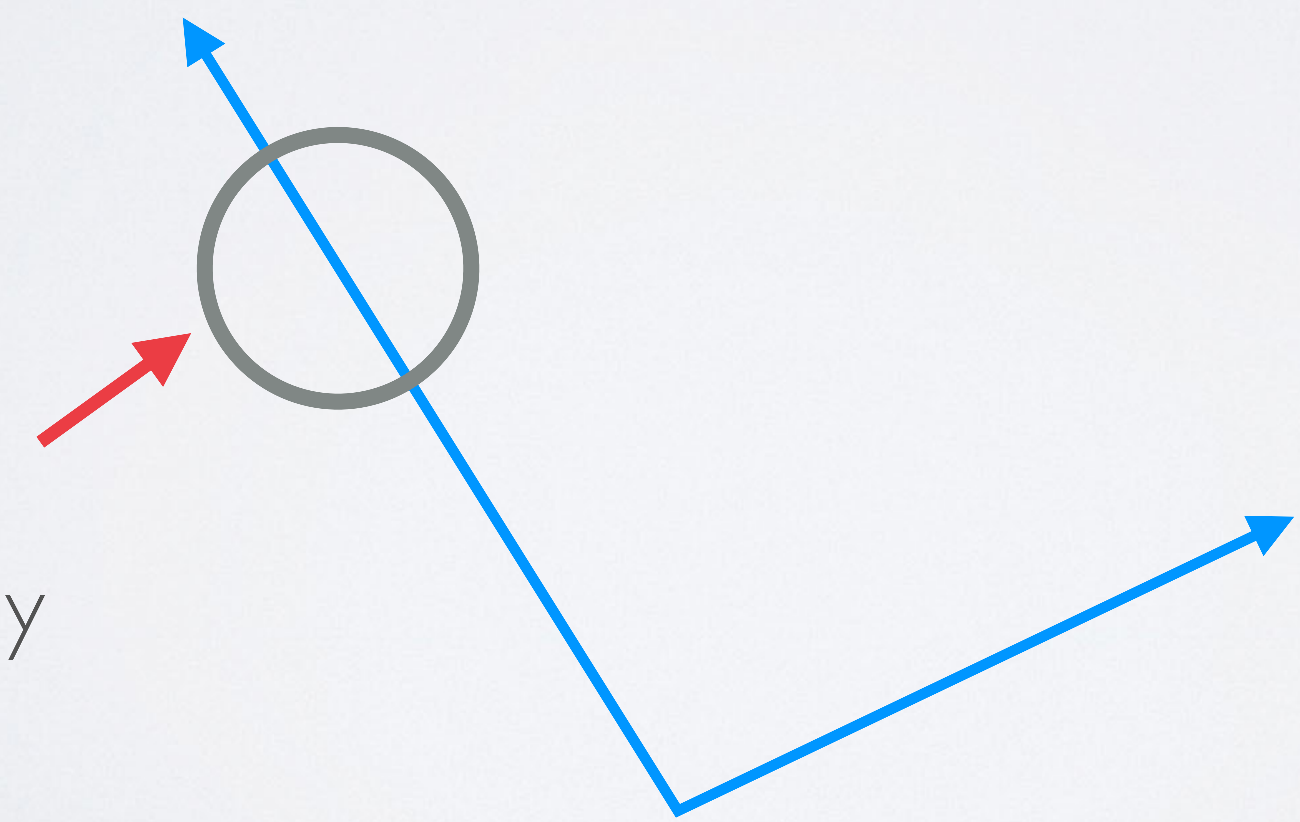


Problem: Newton direction is bad! Excessive backtracking!

why?

GEOMETRIC INTERPRETATION

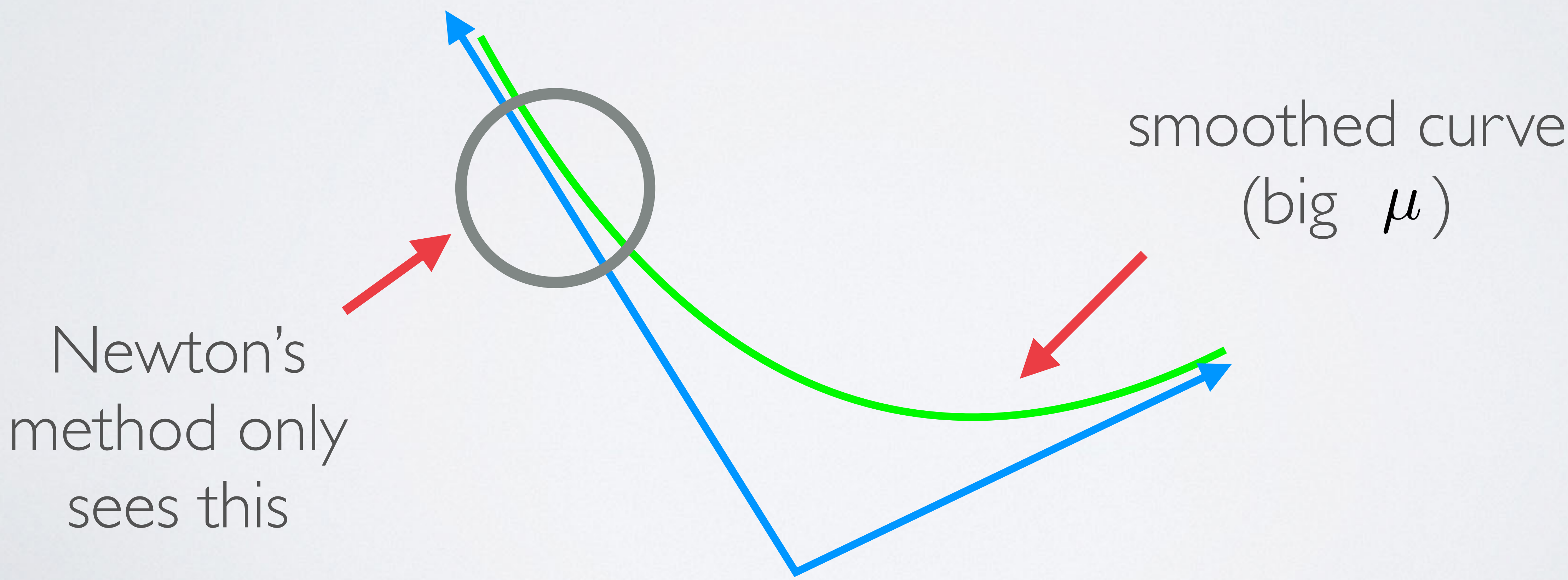
Minimize this?
How far should you go?



Newton's
method only
sees this

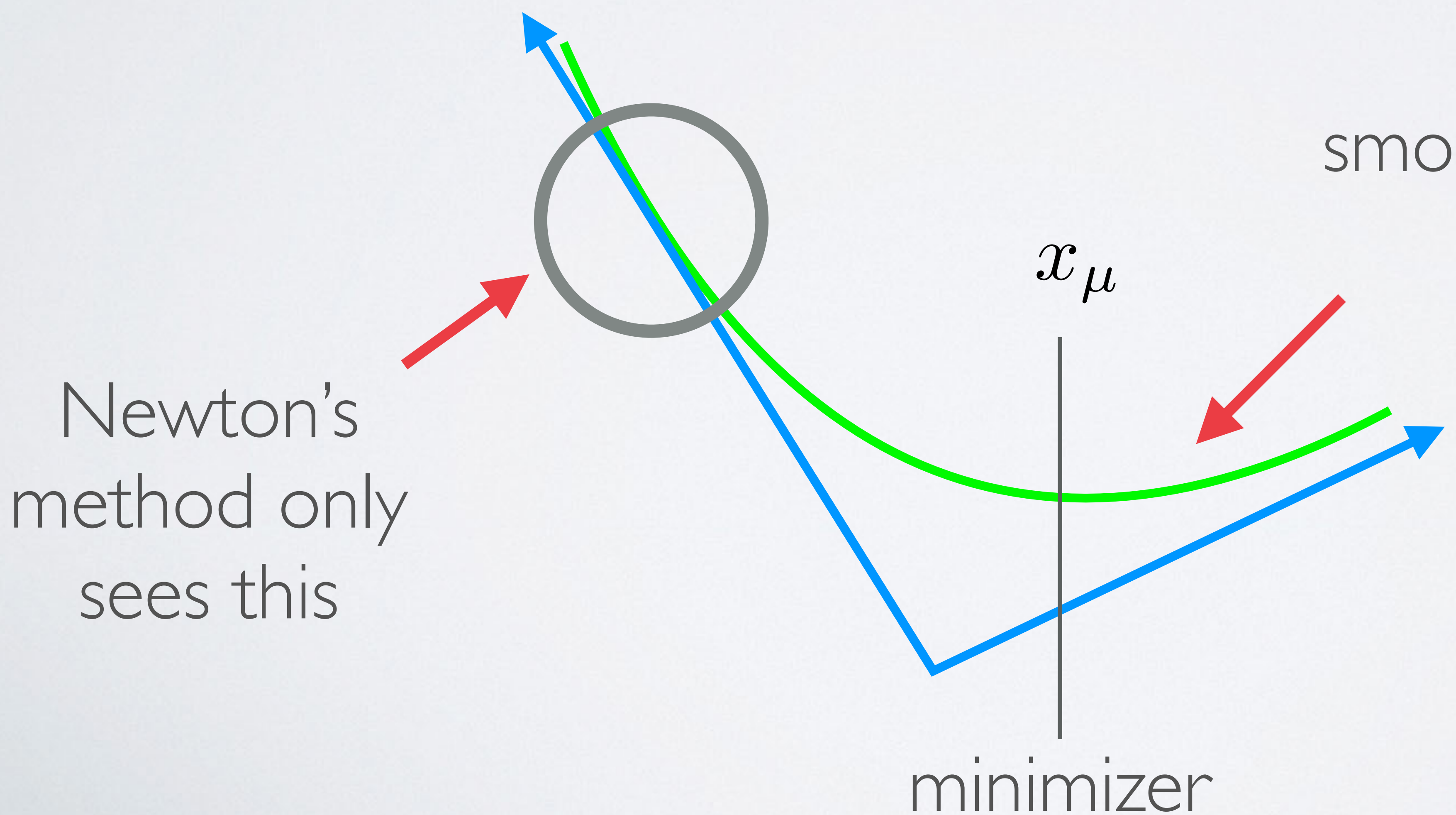
GEOMETRIC INTERPRETATION

Smoothing **globalizes** information



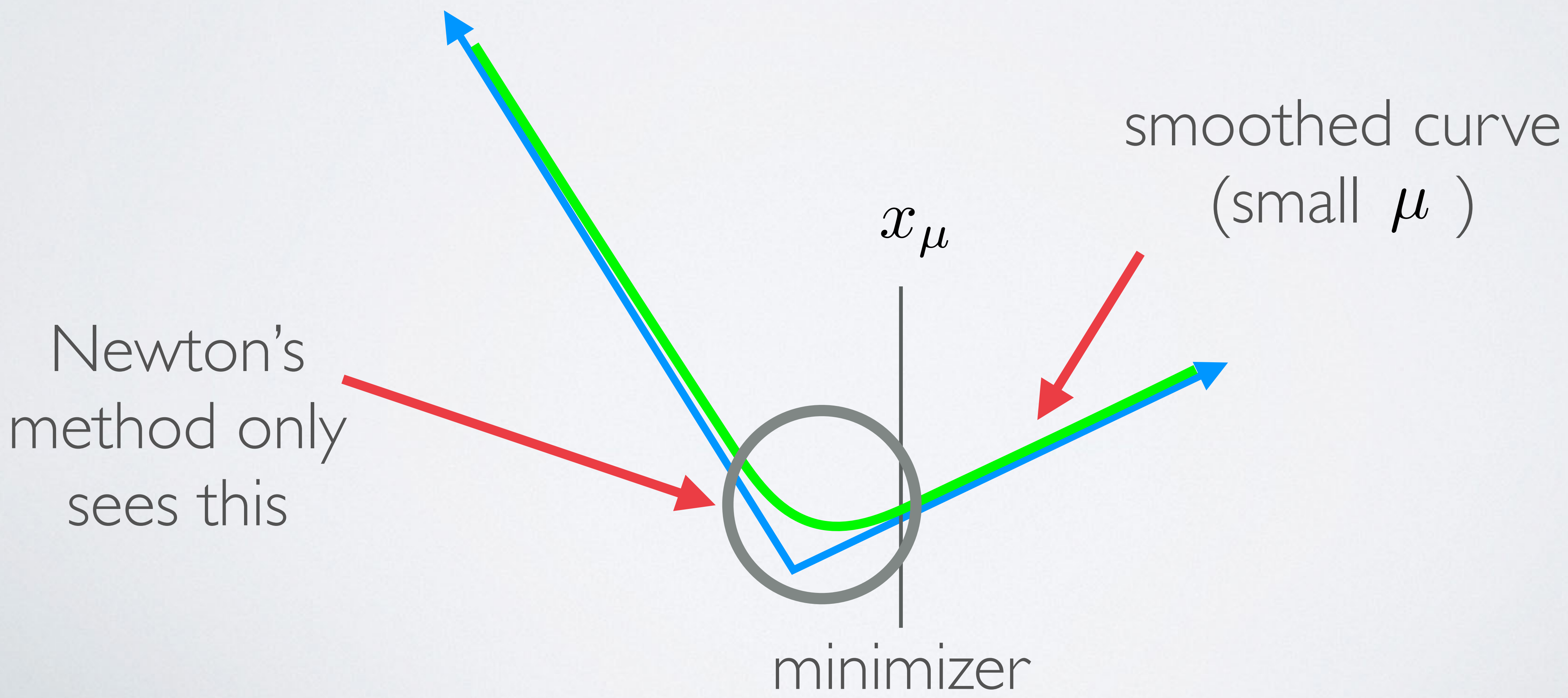
GEOMETRIC INTERPRETATION

Smoothing **globalizes** information



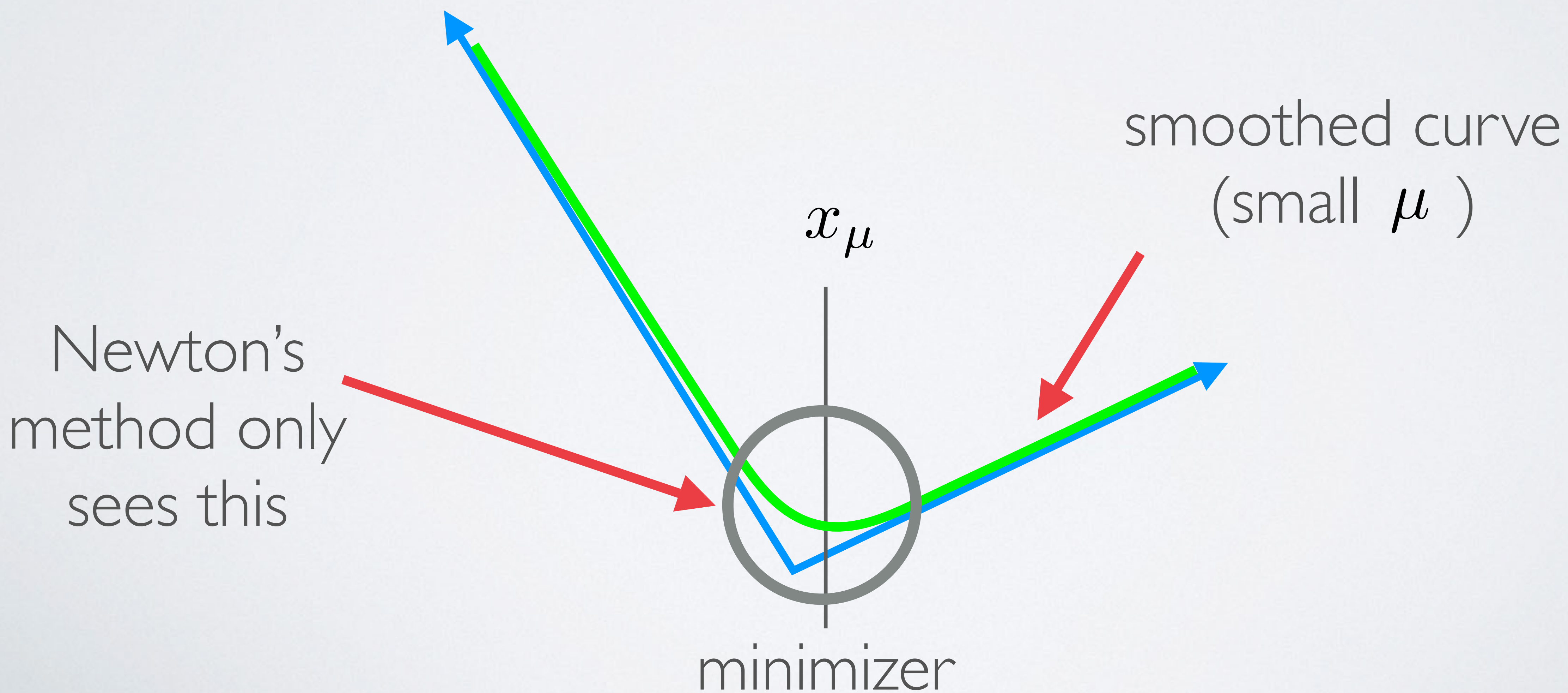
GEOMETRIC INTERPRETATION

Smoothing **globalizes** information



GEOMETRIC INTERPRETATION

Smoothing **globalizes** information



BARRIER IP METHOD

“centering step”

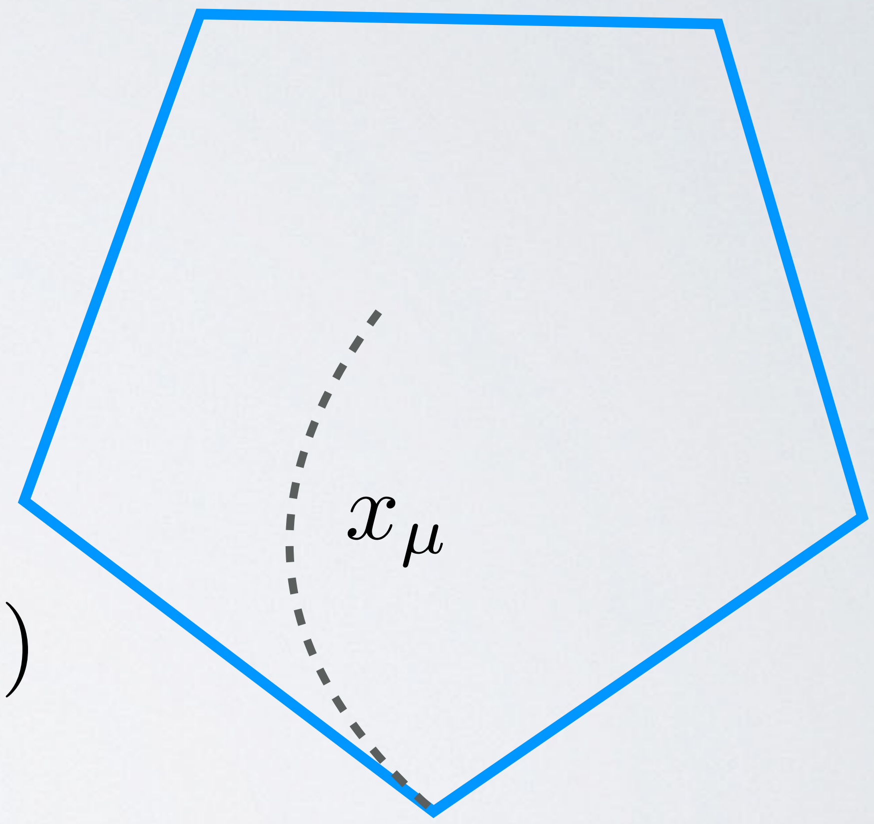
“central path”

While “not converged:”

- Choose some large μ
- Solve for x_μ

$$\begin{array}{ll} \text{minimize} & f_0(x) - \mu \log(-g(x)) \\ \text{subject to} & Ax = b \end{array}$$

- $\mu \leftarrow \mu/10$



idea: collapse the quadratic region around the solution

CENTRAL PATH

$$\begin{aligned} &\text{minimize} \quad f(x) - \sum_{i=1}^{m} \mu \log(-g_i(x)) \\ &\text{subject to} \quad Ax = b \end{aligned}$$

optimality condition

$$\nabla f(x) - \sum_{i=1}^{m} \nabla g_i(x) \frac{\mu}{g_i(x)} + A^T \nu = 0$$

Lagrangian

$$L(x, \lambda, \nu) = f(x) + \langle \nu, g \rangle + \langle \lambda, Ax - b \rangle$$

CENTRAL PATH

optimality condition

$$\nabla f(x) - \sum_{i=1}^{m} \nabla g_i(x) \frac{\mu}{g_i(x)} + A^T \nu = 0$$

Lagrangian

$$L(x, \lambda, \nu) = f(x) + \langle \nu, g \rangle + \langle \lambda, Ax - b \rangle$$

optimality condition

$$\partial_x L = \nabla f(x) + \sum \nabla g_i(x) \nu_i + A^T \lambda = 0$$

Lagrange multipliers $\nu_i = -\frac{\mu}{g_i(x)}$

LAGRANGIAN INTERPRETATION

A central point minimizes the Lagrangian with Lagrange multipliers

$$\nu_i = -\frac{\mu}{g_i(x)}$$

Given x_μ:

primal objective: $f(x_\mu)$

$$\begin{aligned} d(\lambda, \nu) &= f(x_\mu) + \sum_{i=1}^{m} \nu_i g_i(x_\mu) + \langle \lambda, Ax_\mu - b \rangle \\ &= f(x_\mu) - \sum_{i=1}^{m} \frac{\mu}{g_i(x_\mu)} g_i(x_\mu) + \langle \lambda, 0 \rangle \\ &= f(x_\mu) - m\mu \end{aligned}$$

$$\text{duality gap} = f(x_\mu) - d(\lambda, \nu) = m\mu$$

DUALITY GAP

Points on the central path have small duality gap!

$$\text{duality gap} = f(x_\mu) - d(\lambda, \nu) = m\mu$$

This proves that as μ gets small, we approach a solution.

If we shrink by constant factor of 10, then $\text{gap} < \epsilon$ after $\log_{10}(\epsilon)$ iterations!

COMPLEXITY

Interior point methods achieve polynomial complexity for most convex programs, including LP's!

Theorem

The total number of Newton steps needed to get a duality gap less than ϵ is bounded by

path-following

$$\left\lceil \sqrt{m}\log_2\left(\frac{m\mu_0}{\epsilon}\right)\right\rceil \left(\frac{1}{2\gamma} + \log_2 \log_2 \epsilon\right)$$

centering

where γ is a constant that only depends on the backtracking parameters for the Armijo line search.

see Boyd and Vandenberghe, sec 11.5

PRIMAL-DUAL IP'S

standard form LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Lagrangian

$$L(x, \lambda, \nu) = c^T x + \langle \nu, -x \rangle + \langle \lambda, b - Ax \rangle$$

KKT system

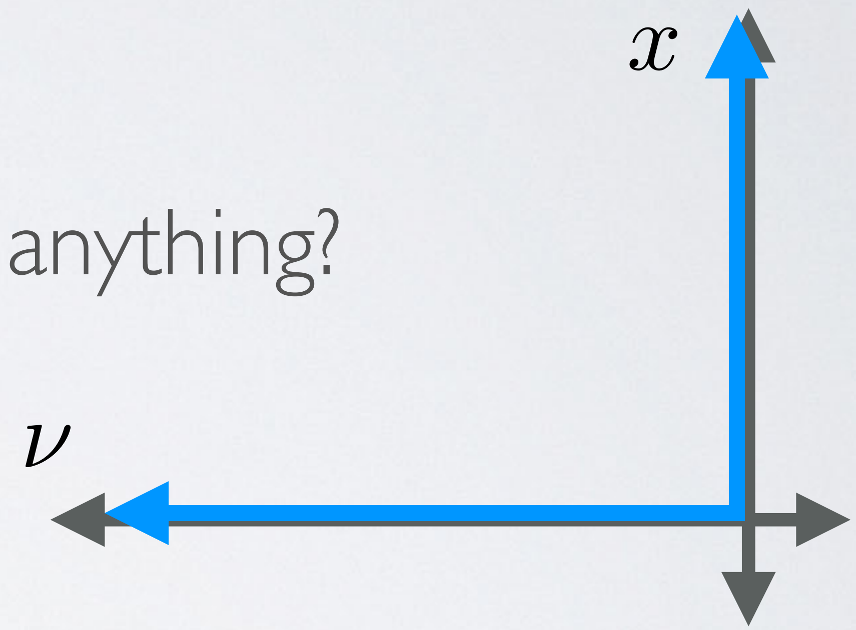
$A^T \lambda + \nu = c$	primal optimality
$Ax = b$	dual optimality
$\nu_i x_i = 0, \ \forall i$	comp slackness
$x, \nu \geq 0$	primal/dual feasibility

PRIMAL-DUAL IP'S

Primal-Dual IP's use Newton's method for non-linear equations to solve the KKT system

problem:
this equation
is non-smooth

remind you of anything?



KKT system

$$A^T\lambda + \nu = c$$ primal optimality

$$Ax = b$$ dual optimality

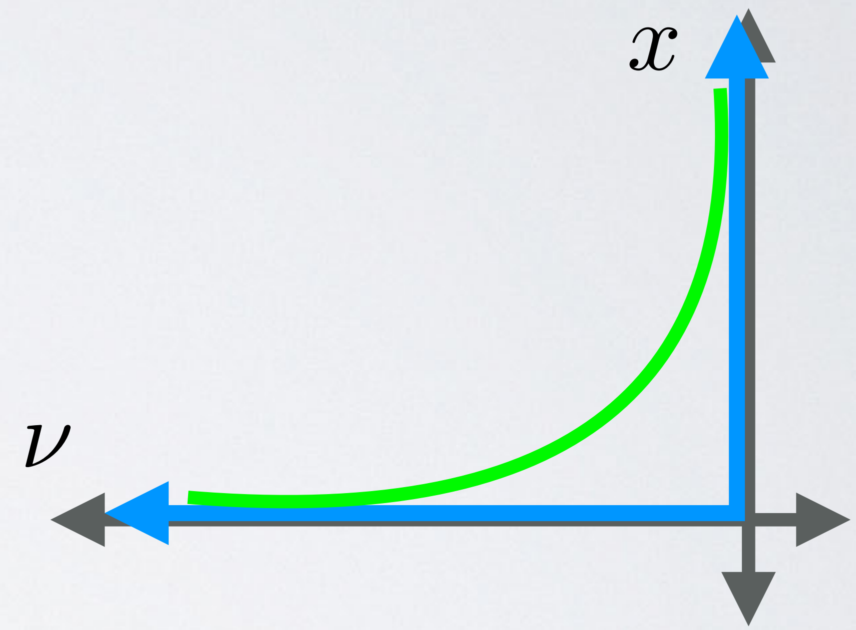
$$\nu_i x_i = 0,\ \forall i$$ comp slackness

$$x, \nu \geq 0$$ primal/dual feasibility

SMOOTHED KKT

$$\nu_i x_i = 0 \longrightarrow \nu_i x_i = \mu$$

$$\text{Let } \mu \to 0$$



KKT system

$A^T\lambda + \nu = c$	primal optimality
$Ax = b$	dual optimality
$\nu_i x_i = \mu, \ \forall i$	comp slackness
$x, \nu \geq 0$	primal/dual feasibility

NEWTON STEP

$$F(x, \lambda, \nu) = \begin{pmatrix} A^T\lambda + \nu - c \\ Ax - b \\ XV - \mu \end{pmatrix}$$

$$X = \mathrm{diag}(x)$$
$$V = \mathrm{diag}(\nu)$$

modified equation

$$X = \mathrm{diag}(x),\ V = \mathrm{diag}(\nu)$$

solve $F(x, \lambda, \nu) = 0$

KKT system

$A^T\lambda + \nu = c$	primal optimality
$Ax = b$	dual optimality
$\nu_i x_i = \mu,\ \forall i$	comp slackness
$x, \nu \geq 0$	primal/dual feasibility

NEWTON STEP

$$F(x,\lambda,\nu) = \begin{pmatrix} A^T\lambda + \nu - c \\ Ax - b \\ XV - \mu \end{pmatrix}$$

$$X = \mathrm{diag}(x)$$
$$V = \mathrm{diag}(\nu)$$

solve $F(x,\lambda,\nu) = 0$

$$J(x,\lambda,\nu)\begin{pmatrix} \Delta x \\ \Delta\lambda \\ \Delta x \end{pmatrix} = -F(x,\lambda,\nu)$$

Newton system

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ V & 0 & X \end{pmatrix}\begin{pmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{pmatrix} = \begin{pmatrix} -r_c \\ -r_b \\ -XV\mathtt{1} + \mu\mathtt{1} \end{pmatrix}$$

PRIMAL DUAL IP METHOD

Secret Sauce

PD methods do not have separate centering steps.
They shoot straight for the solution

While "not converged:"

- Calculate average "duality measure:" $\mu = \frac{x^T\nu}{n}$
- Calculate target duality measure: $\sigma\mu$

agressiveness

- Solve the Newton system

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ V & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta\lambda \\ \Delta\nu \end{pmatrix} = \begin{pmatrix} -r_c \\ -r_b \\ -XV\mathbf{1} + \sigma\mu\mathbf{1} \end{pmatrix}$$

- Update iterates

$$x \leftarrow x + \alpha_p \Delta x$$

$$(\lambda, \nu) \leftarrow (\lambda, \nu) + \alpha_d(\Delta\lambda, \Delta\nu)$$

stepsize
chosen to stay
near central path

ADAPTIVE PARAMETER CHOICE

$$x \leftarrow x + \alpha_p \Delta x$$

$$(\lambda, \nu) \leftarrow (\lambda, \nu) + \alpha_d (\Delta\lambda, \Delta\nu)$$

find largest steps that don't violate feasibility $\quad x, \nu \geq 0$

$$\alpha_p^{max} = \min_{\Delta x_i < 0} \frac{x_i}{\Delta x_i} \qquad \alpha_d^{max} = \min_{\Delta \nu_i < 0} \frac{\nu_i}{\Delta \nu_i}$$

$$\alpha_p = \min\{1, \eta\alpha_p^{max}\} \qquad \alpha_d = \min\{1, \eta\alpha_d^{max}\}$$

$$\mu_{target} = \sigma\mu$$

$$\sigma = \left(\frac{(x + \alpha_p \Delta x)^T (\nu + \alpha_d \Delta \nu)}{\mu} \right)^3$$

new duality

old duality

see Necedal and Wright, chapter 14

COMPOSITE NEWTON'S METHOD

classical Newton

factorizing this is expensive!

$$J(x)\Delta x = -F(x)$$

$$x \leftarrow x + \alpha \Delta x$$

Let's try to use it twice!

predictor $$J(x)\Delta x_p = -F(x)$$

corrector $$J(x)\Delta x_c = -F(x + \Delta x_p)$$

update $$x \leftarrow x + \alpha(\Delta x_p + \Delta x_c)$$

Why would you do this?

COMPOSITE NEWTON'S METHOD

predictor $$J(x)\Delta x_p = -F(x)$$

corrector $$J(x)\Delta x_c = -F(x+\Delta x_p)$$

update $$x \leftarrow x + \alpha(\Delta x_p + \Delta x_c)$$

Allows you to use the Hessian 2X!

Can prove **cubic** convergence (vs **quadratic**)

better than Newton? why?

It's **free**!

see Tapia, Zhang, Saltzman, Weiser '90

MEHROTRA'S PREDICTOR-CORRECTOR METHOD

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ V & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \\ \Delta \nu \end{pmatrix} = \begin{pmatrix} -r_c \\ -r_b \\ -XV\mathbf{1} \end{pmatrix}$$

subtract from here

$$(x + \Delta x)^T (\nu + \Delta \nu) = \underbrace{x^T \nu + x^T \Delta \nu + \nu^T \Delta x}_{=0} + \Delta x^T \Delta \nu = \Delta x^T \Delta \nu$$

I wanted this to be 0!

MEHROTRA'S METHOD

predictor

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ V & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x_p \\ \Delta \lambda_p \\ \Delta \nu_p \end{pmatrix} = \begin{pmatrix} -r_c \\ -r_b \\ -XV\mathsf{1} \end{pmatrix}$$

we now have

$$(x + \Delta x_p)^T (\nu + \Delta \nu_p) = \Delta x_p^T \Delta \nu_p$$

corrector

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ V & 0 & X \end{pmatrix} \begin{pmatrix} \Delta x_c \\ \Delta \lambda_c \\ \Delta \nu_c \end{pmatrix} = \begin{pmatrix} -r_c \\ -r_b \\ -XV\mathsf{1} + \sigma\mu\mathsf{1} - \Delta X_p \Delta V_p \end{pmatrix}$$

finally

$$x \leftarrow x + \alpha_p \Delta x_c$$

$$(\lambda, \nu) \leftarrow (\lambda, \nu) + \alpha_d (\Delta \lambda_c, \Delta \nu_c)$$

PHASE I VS PHASE II

IP's require **strict** feasibility

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \\ & g(x) \leq 0 \end{aligned}$$

starting point satisfies

$$Ax = b$$

$$g(x) < 0$$

strict

phase I

$$\begin{aligned} \text{minimize} \quad & z \\ \text{subject to} \quad & Ax = b \\ & g_i(x) \leq z, \ \forall i \end{aligned}$$

$\xrightarrow{\text{if } z<0}$

phase II

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \\ & g(x) \leq 0 \end{aligned}$$

MOST PROBLEMS HAPPEN IN PHASE I

redundant constraints

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \end{aligned} \longrightarrow \begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax \leq b \\ & Ax \geq b \end{aligned}$$

will this work?

(very) poor conditioning

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \\ & g(x) \leq 0 \end{aligned}$$

difficult to invert

PROPERTIES OF IP METHODS

pro's

fast linear/superlinear convergence

very high precision achievable

no asymptotic slowdown

extremely reliable

totally automated

con's

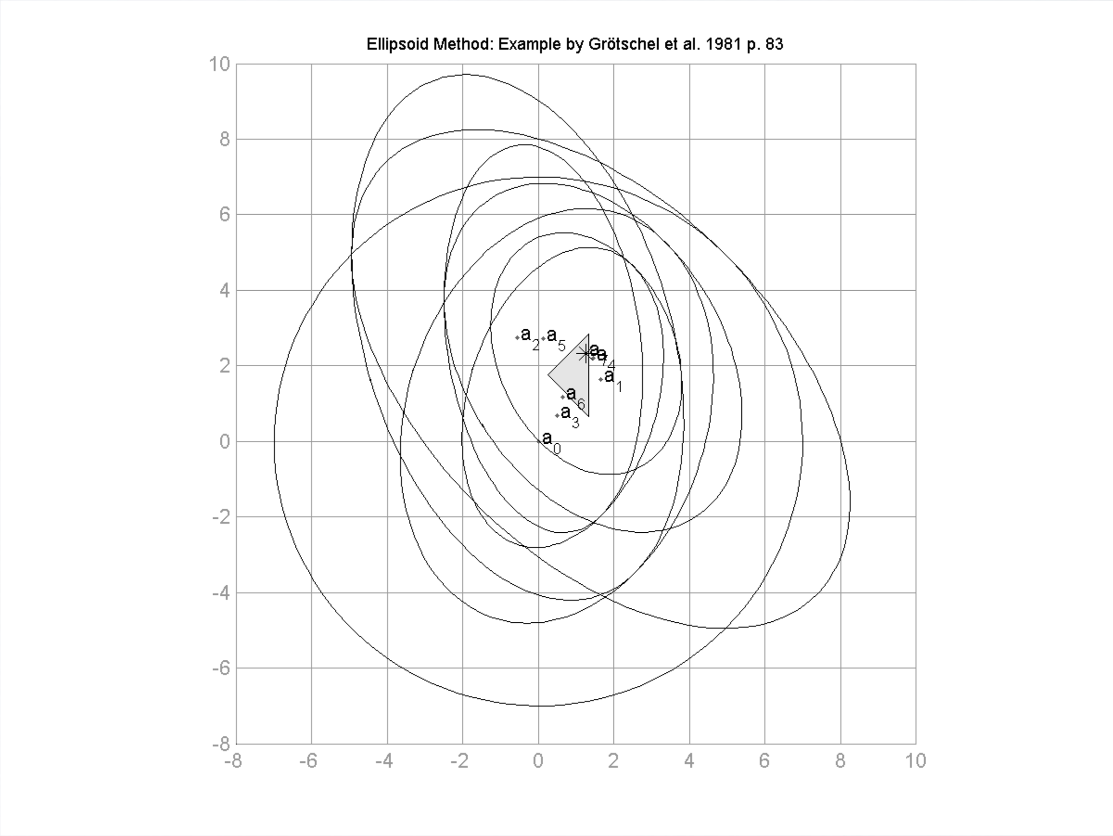
high per-iteration complexity

useless for more than 10K unknowns

COMPLEXITY

1979 **Ellipsoid method : Leonid Khachiyan**

First polynomial-time proof for LP method



Solves LP's with n variables and b bits of precision in $O(N^6L)$ time

Solves most standard convex problems in poly time

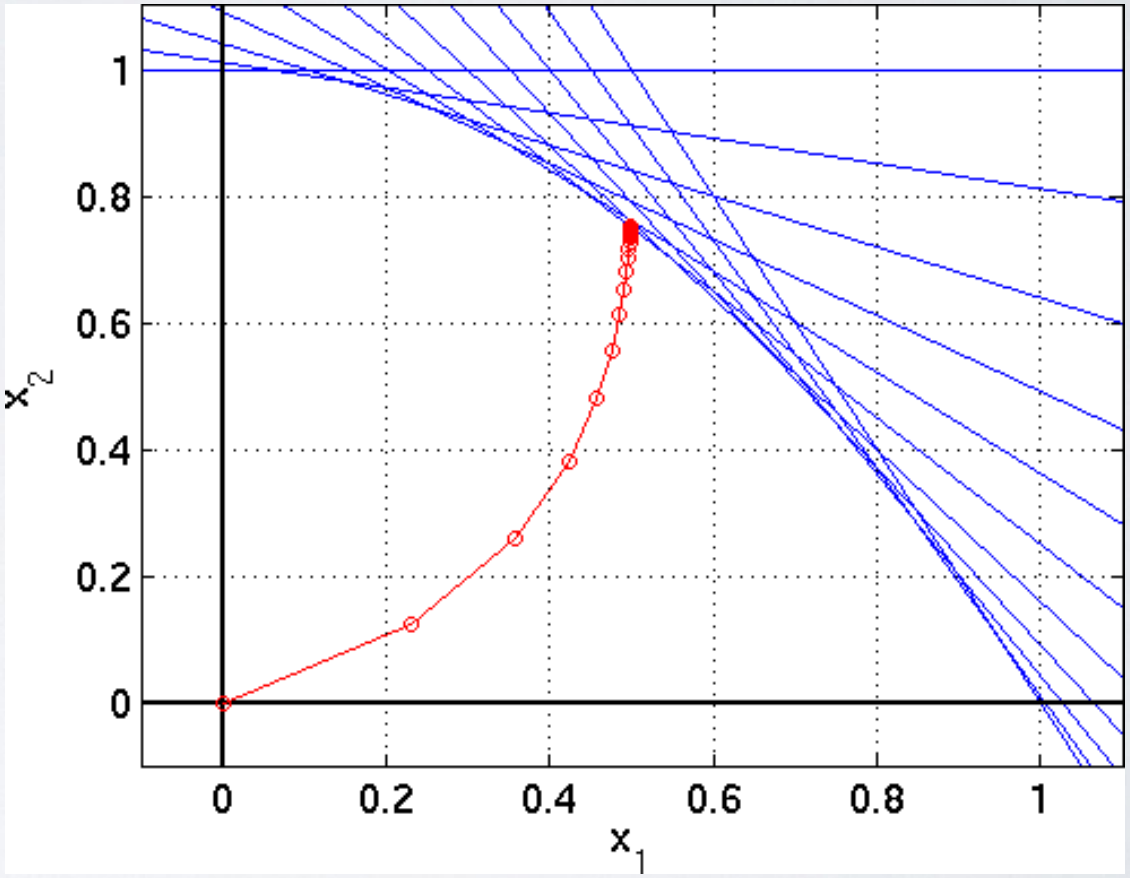
COMPLEXITY

1984 **Projection Algorithm : Narendra Karmarkar**

First polynomial-time proof for *efficient* LP method

Solves LP's with n variables and b bits of precision in $O(N^{3.5}L)$ with smaller constant

"Faster than barrier method," but Philip Gill proved they are the same



Other IP methods proved polynomial: Yurii Nesterov and others

BIG OPEN PROBLEMS

s there a strongly polynomial time algorithm for LP?

Strongly polynomial: runtime independent of input size

ex: sorting

Weakly polynomial: number of operations depends on size of input numbers, not just on how many numbers are input

ex: Euclidean GCD algorithm

Does a polytope with N faces and D dimensions have polynomial diameter?

Diameter: maximum distance/edges connecting 2 vertices

If not then impossible to prove polynomial runtime for simplex method