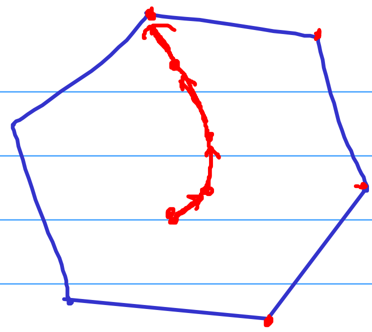
April 19

Interior point Methods



Karmarkar 1984.

- Provably polynomial time
- Beats Simplex in many cases
- also solves convex programs. under certain conditions.
- Need the objective function and constraints explicitly.

Note that some convex programs can be NP-hard solve.

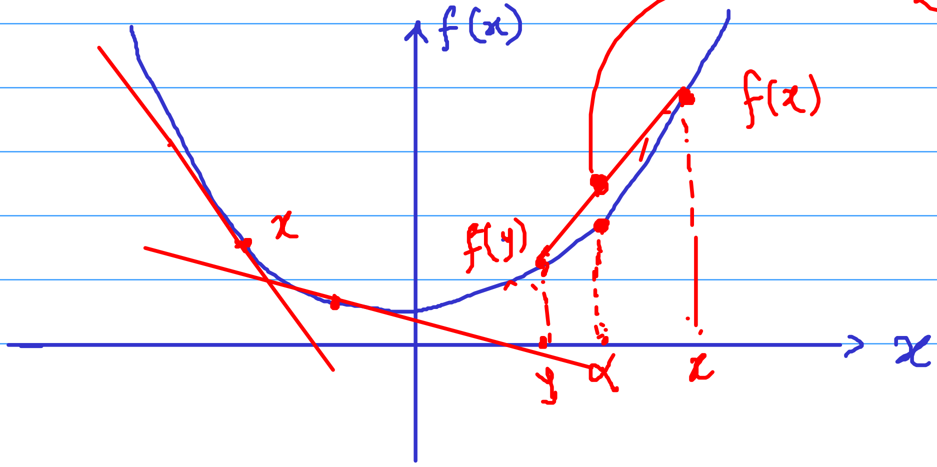
Def (convex functions)

For a convex set $K \subseteq \mathbb{R}^n$, a function $f: K \to \mathbb{R}$ is called convex. if

for any $x, y \in K$ and $0 \leq \lambda \leq 1$

$$\underbrace{f(\lambda x + (1-\lambda) y)}_{\alpha = \frac{x+y}{2}} \leq \lambda f(x) + (1-\lambda) f(y)$$

$$\frac{f(x)+f(y)}{2}$$



Examples

① $x_1^2 + x_2^2 - x_1 x_2$

② $x_1 + x_2 - 3x_3$

③ $-\sin x \quad 0 \leq x \leq \pi$

Univariate

Convex function $\qquad f(y) \geq \underbrace{f(x) + f'(x)(y-x)}_{\text{linear approx at } x}$

for all $y \in K$

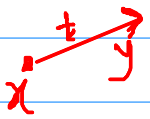
Multivariate

Def (Gradient)

For a differentiable function $f: K \to \mathbb{R}$

$\nabla f : K \to \mathbb{R}$

$$\nabla f(x) = \begin{bmatrix} \partial f / \partial x_1 (x) \\ \partial f / \partial x_2 (x) \\ \partial f / \partial x_3 (x) \\ \vdots \\ \partial f / \partial x_n (x) \end{bmatrix}$$



Fact: Derivative at x in a direction y

$t \in \mathbb{R}$ $\qquad \dfrac{d}{dt} f(x + ty) = \langle \nabla f(x), y \rangle$

Convexity in terms of Gradient

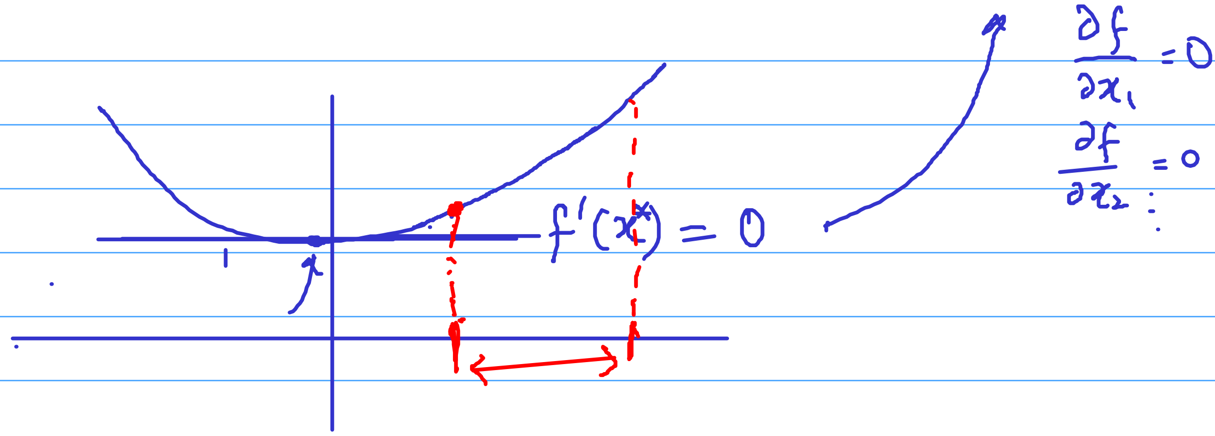
For a differentiable function $f: K \to \mathbb{R}$

f is convex if and only if

$$f(y) \geq \underbrace{f(x) + \langle y - x, \nabla f(x) \rangle}_{\text{linear approx of } f(y) \text{ around } x.}$$

Minimizing a convex function $f: \mathbb{R}^n \to \mathbb{R}$ over $\mathbb{R}^n$ (unconstrained)

x^* is the minimizer for $f(x)$ iff $\nabla f(x^*) = 0$



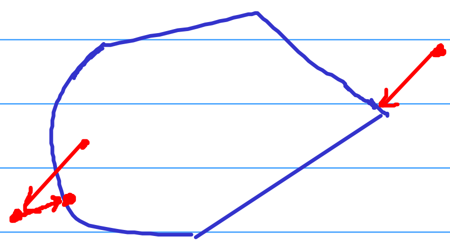
Gradient descent

move in the direction of $-\nabla f(x)$

$$x_{t+1} = x_t - \gamma_t \nabla f(x_t)$$

What about constrained convex minimization?

Projected Gradient descent



Works only if a procedure for projection to K is known.

Projection is Another convex optimization problem.

Some variants (MWU) work but does not solve arbitrary LPs in P.

Newton-Raphson

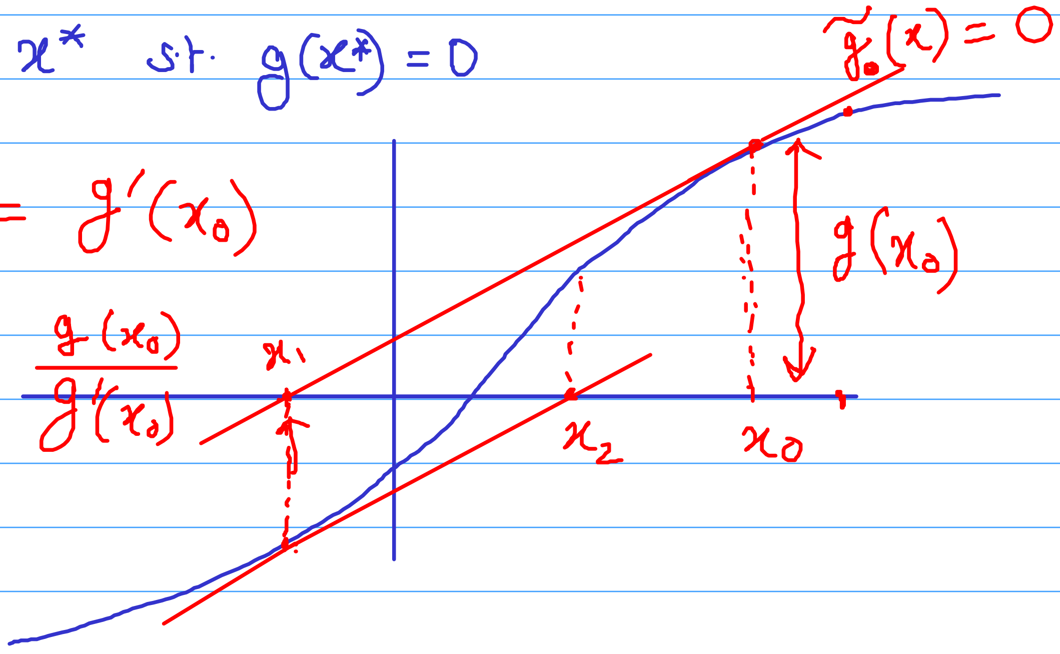
Ref: Vishnoi's book.

Newton's root finding

$g(x) = f'(x)$

Univariate.

find x^* s.t. $g(x^*) = 0$



$$\frac{0 - g(x_0)}{x_1 - x_0} = g'(x_0)$$

$$x_1 = x_0 - \frac{g(x_0)}{g'(x_0)}$$

$$x_{t+1} = x_t - \frac{g(x_t)}{g'(x_t)}$$

$$x_{t+1} = x_t - \frac{f'(x_t)}{f''(x_t)}$$

Faster convergence than gradient descent but need second derivative.

Multivariate

$f : \mathbb{R}^n \to \mathbb{R}$

minimize f

$\nabla f(x) = 0$

$$\frac{\partial f}{\partial x_i} = 0 \quad \forall i$$

n simultaneous equations.

Let $\quad g_i = \dfrac{\partial f}{\partial x_i}$

$g_1(x) = 0$ $\qquad$ n variables

$g_2(x) = 0$ $\qquad$ n equations.

$\vdots$ $\qquad$ $g : \mathbb{R}^n \to \mathbb{R}^n$

$g_n(x) = 0$ $\qquad$ $h(x)$ $\qquad$ $\underbrace{h(x_0) + h'(x_0)(x - x_0)}$

At point x_t, linear approximations of g_i

for all i, $\quad g_i(x_t) + \langle \nabla g_i(x_t), x - x_t \rangle = 0$

$\forall i, \quad \langle \nabla g_i(x_t), x - x_t \rangle = - g_i(x_t)$

$$\begin{bmatrix} \cdots & \nabla g_1(x_t) & \cdots \\ \cdots & \nabla g_2 & \cdots \\ & & \\ \cdots & \nabla g_n & \cdots \end{bmatrix} (x - x_t) = - \begin{pmatrix} g_1(x_t) \\ g_2(x_t) \\ \vdots \\ g_n(x_t) \end{pmatrix} =$$

$$\begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1 \partial x_i} & \dfrac{\partial^2 f}{\partial x_j \partial x_i}(x_t) \\ & \end{bmatrix} (x - x_t) = - \begin{pmatrix} \partial f / \partial x_1 \\ \vdots \\ \partial f / \partial x_n \end{pmatrix} = -\nabla f(x_t)$$

Hessian

$$\nabla^2 f(x_t)(x - x_t) = - \nabla f(x_t)$$

$$x_{t+1} = x_t - \left[\nabla^2 f(x_t)\right]^{-1} \nabla f(x_t)$$

Is $\nabla^2 f$ always invertible?

If we assume that f is strictly convex then it is.

Def (Strictly Convex)

Defined by following equivalent criteria

① for all x, y
for $\lambda \in (0,1)$

$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$$

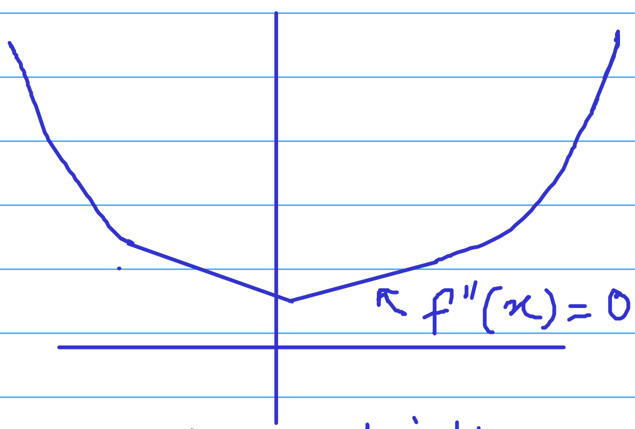
② for all x, y s.t. $x \neq y$

$$f(y) > f(x) + \langle \nabla f(x), y - x \rangle$$

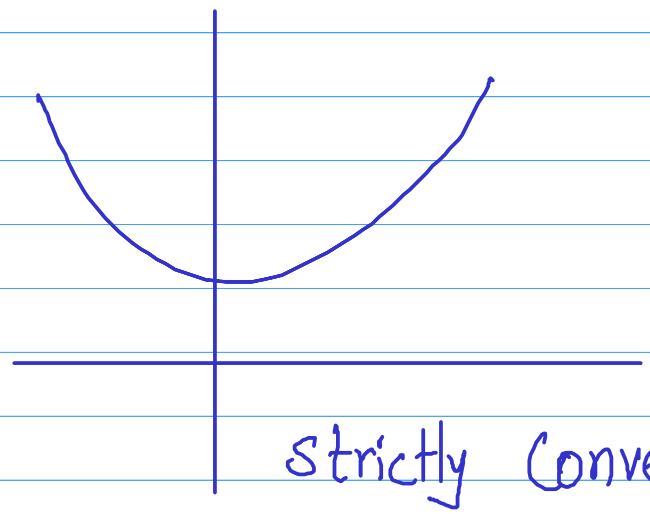
③ for all x

$$\nabla^2 f(x) \succ 0 \quad \text{positive definite}$$

Positive definite matrix is always invertible.



Not strictly convex
$\nabla^2 f$ non-invertible.



Strictly Convex
$\nabla^2 f$ invertible.

Convergence Guarantee in Newton's Method

$f \leftarrow$ strictly convex. Let $H = \nabla^2 f$

NL condition (Newton Local)

For any $\delta < 1$, $\forall$ x, y s.t.

$\|y-x\|_x$ $\quad (y-x)^T H(x)(y-x) \leq \delta$

we have

$$(1-3\delta)H(x) \preceq H(y) \preceq (1+3\delta)H(x)$$

Guarantee

Let f satisfy NL condition

Let x_0 be s.t. $\left|\nabla f(x_0)^T H(x_0)^{-1} \nabla f(x_0)\right| \leq \frac{1}{16}$

and $x_1 = x_0 - \left[\nabla^2 f(x_0)\right]^{-1} \nabla f(x_0)$

then

$$\left|\nabla f(x_1)^T H(x_1)^{-1} \nabla f(x_1)\right| \leq 3 \left|\nabla f(x_0)^T H(x_0)^{-1} \nabla f(x_0)\right|^2$$

If we want ε-accuracy, i.e, want a point x s.t. $|f(x) - f(x^*)| \leq \varepsilon$ then

Newton's Method no. of iterations $\propto \log(1/\varepsilon)$

Gradient Descent no. of iterations $\propto 1/\varepsilon$

Moving from Unconstrained to Constrained.

LP

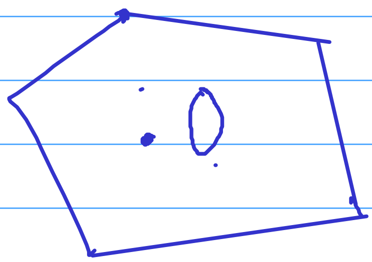
$$\begin{aligned} \min\ & w^T x \\ \text{s.t.}\ & a_1^T x \geq b_1 \\ & a_2^T x \geq b_2 \\ & \vdots \\ & a_m^T x \geq b_m \end{aligned}$$

Introduce a barrier function in the objective

New objective

$$\min\ \eta w^T x + F(x) \quad \text{over } x \in \mathbb{R}^n$$

$P =$



∞

Perfect F

$$\begin{aligned} F(x) &= 0 && \text{if } x \in P \\ &= \infty && \text{if } x \notin P \end{aligned}$$

Not really helpful.

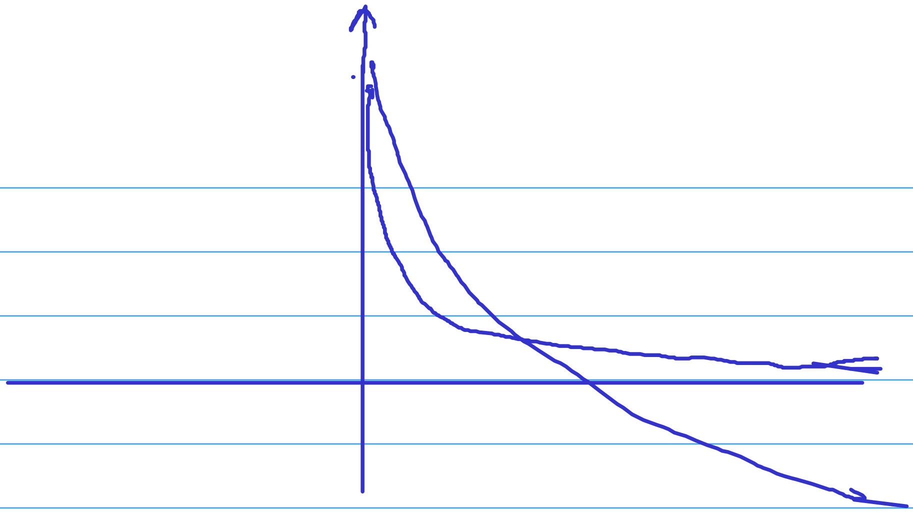
Requirements on barrier $F(x)$

1. Strictly convex
2. should go to infinity as we go towards the boundary.
3. should be decreasing.

x

$x \geq 0$

$-\log x = \log \frac{1}{x}$



Candidates. $-\log x = \log \frac{1}{x}$

$1/x \qquad e^{1/x}$

Right Choice turns out to be $-\log$ $\qquad \underbrace{a_j^T x - b \geq 0}$

$$f_\eta = \eta \underbrace{\omega^T x} + \sum_{j=1}^{m} -\log\left(a_j^T x - b\right)$$

Min f_η

If η is small feasibility is enforced but we will always be away from the boundary. And x^* is on the boundary

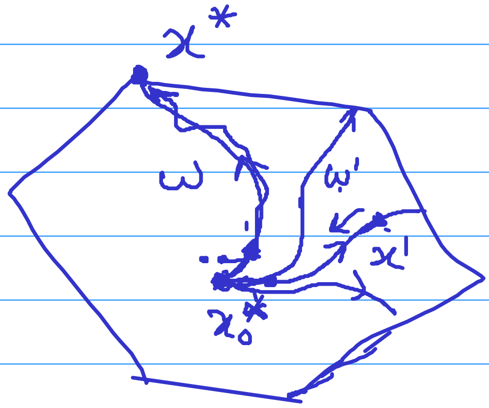
As $\underbrace{\eta \to \infty}$ $\quad x_\eta^* \longrightarrow x^*$

x^*: min of $\omega^T x$

x_η^*: minimizer of f_η

Central Path:

$$\{ x_\eta^* : \eta \geq 0 \}$$



x_0^* minimizer of $F(x)$

Assume that we have

x_0 which is close enough to x_0^*

$x_0 \to x_1 \to x_2 \; \cdots \; \to x_T$

$\eta = \frac{1}{2m}$ $\quad \eta_1 \quad \eta_2 \quad \cdots \quad \eta_T$

Maintain $\quad x_i$ is close to $x_{\eta_i}^*$

Initial $\quad \eta_0 \approx 1/2m$

x_0 is close to $x_{\eta_0}^* \approx x_0^*$

Update $\quad x_t, \eta_t$

Newton's step $\quad x_{t+1} = x_t - H_{\eta_t}(x_t) \nabla f_{\eta_t}(x_t)$

x_{t+1} is further close $x_{\eta_t}^*$

Update $\quad \eta_{t+1} \leftarrow \eta_t \left(1 + \frac{1}{\sqrt{m}}\right)$

x_{t+1} is close to $x_{\eta_{t+1}}^*$

Termination

$$\eta \approx \frac{m}{\varepsilon}$$

$$\varepsilon = \left| w^T(x^*) - w^T(x_T) \right|$$

How to choose x_0 ?

$x' \in \underbrace{\mathrm{Int}(P)}$.

$$f_\eta = \eta \cdot w^T x + F(x)$$

$$f'_\eta = \eta \cdot w'^T x + F(x)$$

Find w' s.t. x' is on the central path corr to f'_η.

$$x' = x^{*'}_\eta \quad \text{minimizer of } f'_\eta$$

$$\nabla f'_\eta = 0$$

$$\eta w' + \nabla F = 0$$

$$\eta = 1 \qquad w' = -\nabla F$$

Convergence Guarantee in Gradient Descent

Unconstrained.

$f \leftarrow$ convex

$$x_{t+1} \leftarrow x_t - \eta \nabla f(x_t)$$

Initial Point x_0

Parameters : (1) $D = \max \{ \|x - x^*\| : f(x) \leq f(x_0) \}$

(2) $L = \max \dfrac{\|\nabla f(y) - \nabla f(x)\|}{\|y - x\|}$

inside the $B(x^*, D)$

(3) ε : desired closeness

$$\varepsilon = |f(x^*) - f(x_T)|$$

Then the required number of iterations

$$T = O\left(\frac{LD^2}{\varepsilon}\right).$$