Optimization problems

$\min \ f_0(x)$

s.t. $f_i(x) \leq 0 \quad i = 1, \dots, m$

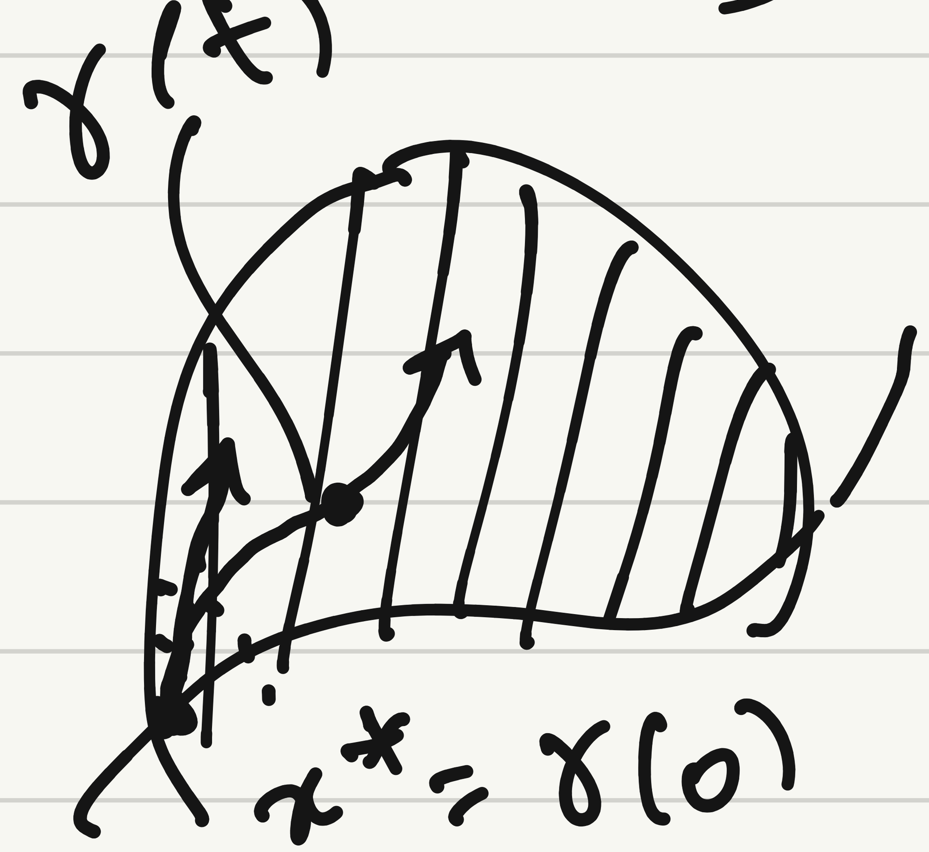
$h_i(x) = 0 \quad i = 1, \dots, p$

Convex if $f_0, f_1, \dots$ convex and h_i are affine

find locally optimal sol. x^* : $f(x^*) \leq f(x)$ for all nearby feasible x.

Recognizing local minima

Assuming f_0 is differentiable.



$\gamma : [0, \dots) \to \mathbb{R}^n, \quad \gamma(0) = x^*$

$\hat{f}_0(t) = f(\gamma(t)), \quad \gamma'(0) = s$: $\boxed{\text{feasible direction}}$

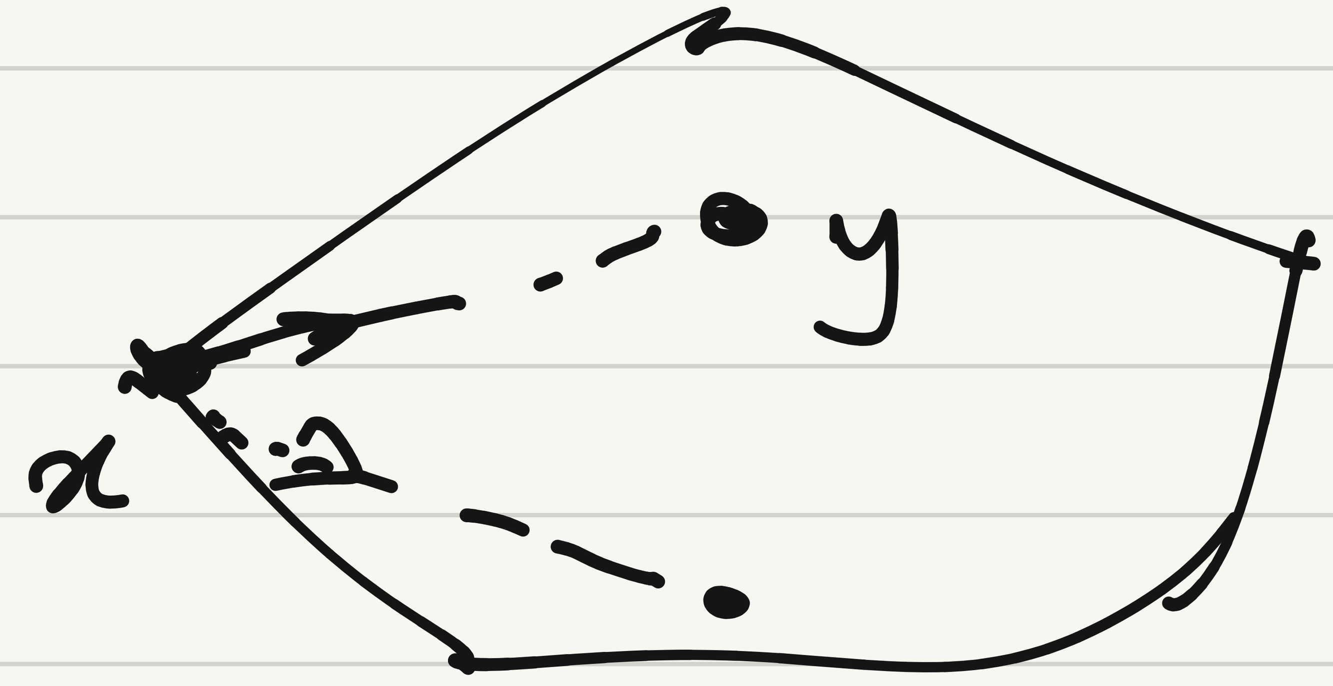
$\hat{f}'(0) = \boxed{\nabla f(x^*)^T s \geq 0}$

$\nabla f(x^*)^T s \geq 0$ for all feasible directions s : $\boxed{\text{first-order necessary condition}}$

Unconstrained case: all directions are feasible! $\Rightarrow$ $\nabla f(x^*) = 0$

Convex case: all feasible dirs. are multiples of $y - x$ for some feasible y

$\nabla f(x)^T (y - x) \geq 0$ for all feas. y.



Convex: necc. conds are also sufficient!

Nonconvex uncons.: $\nabla f(x) = 0$ and $\nabla^2 f(x) \succeq 0$: 2nd order necc. cond.

" " $\nabla^2 f(x) \succ 0$: " " suff. cond.

$f(x) = x^4$, $f(x) = -x^4$

Equivalent opt. probs:

- Change of var. $\min_x f(x) \longleftrightarrow \min_z f(\phi(z))$

  $x = \phi(z)$

- Transformation of fn.: $\min f(x) \longleftrightarrow \min \psi(f(x))$

$\|Ax-b\|_2$

$\to \|Ax-b\|_2^2$

- $f_i(x) \le 0 \longleftrightarrow \overbrace{f_i(x) + \underset{\boxed{\text{slack variable}}}{s} = 0}$ and $\underline{s \ge 0}$

Unconstrained Minimization

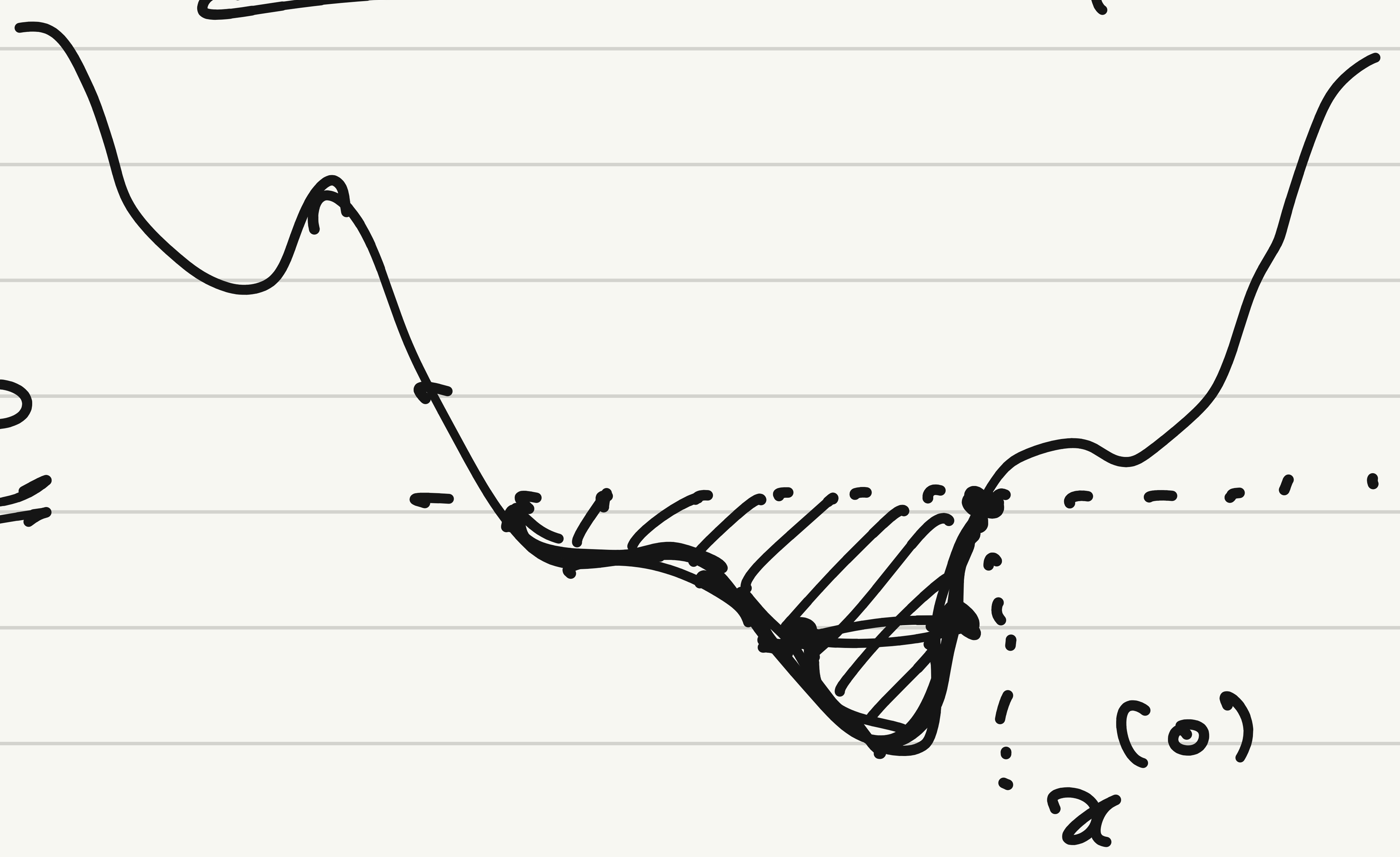
min $f(x)$ over all $x \in \mathbb{R}^n$, assume f is twice differentiable

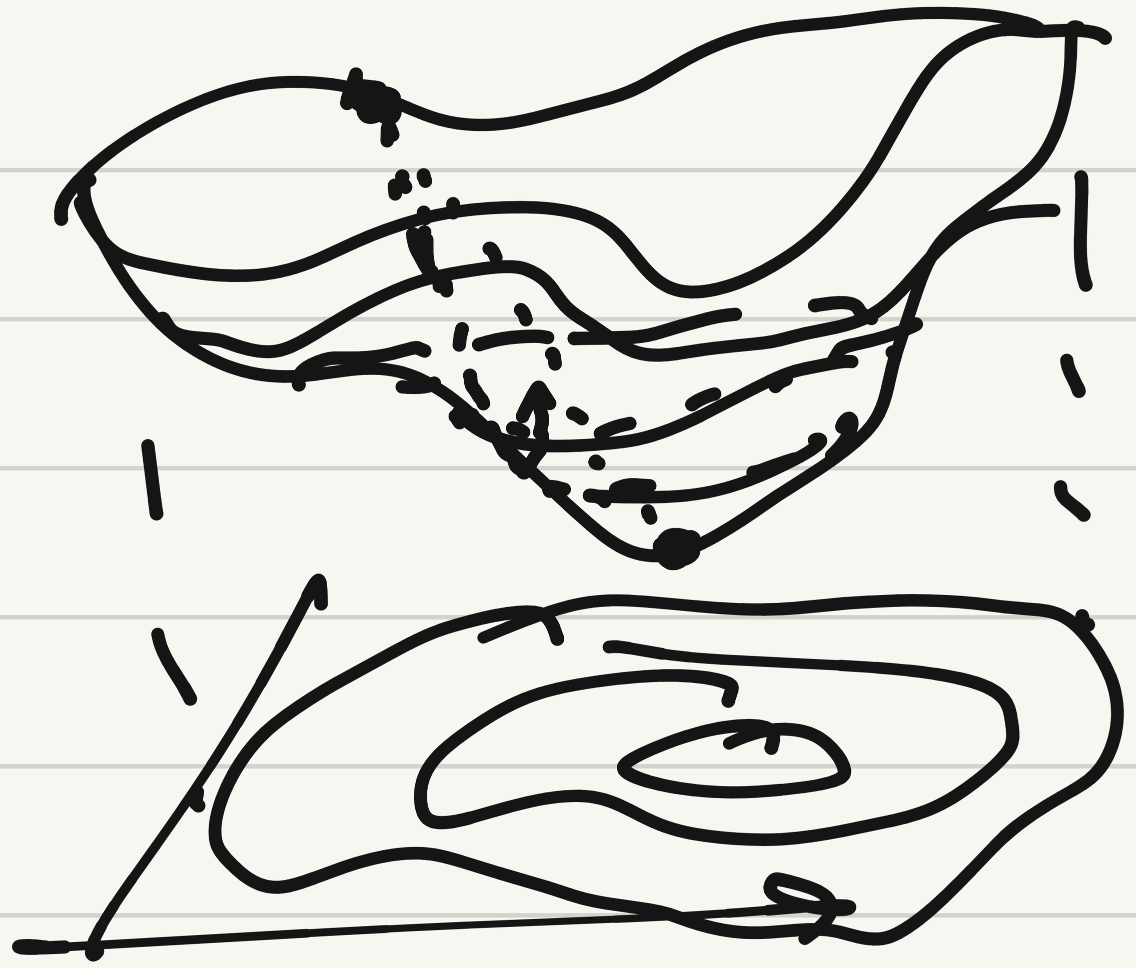
find x^* s.t. $\nabla f(x^*) = 0$

Initial guess $x^{(0)} \rightarrow x^{(1)} \rightarrow x^{(2)} \rightarrow \cdots \rightarrow x^*$: $\boxed{\text{minimizing sequence}}$

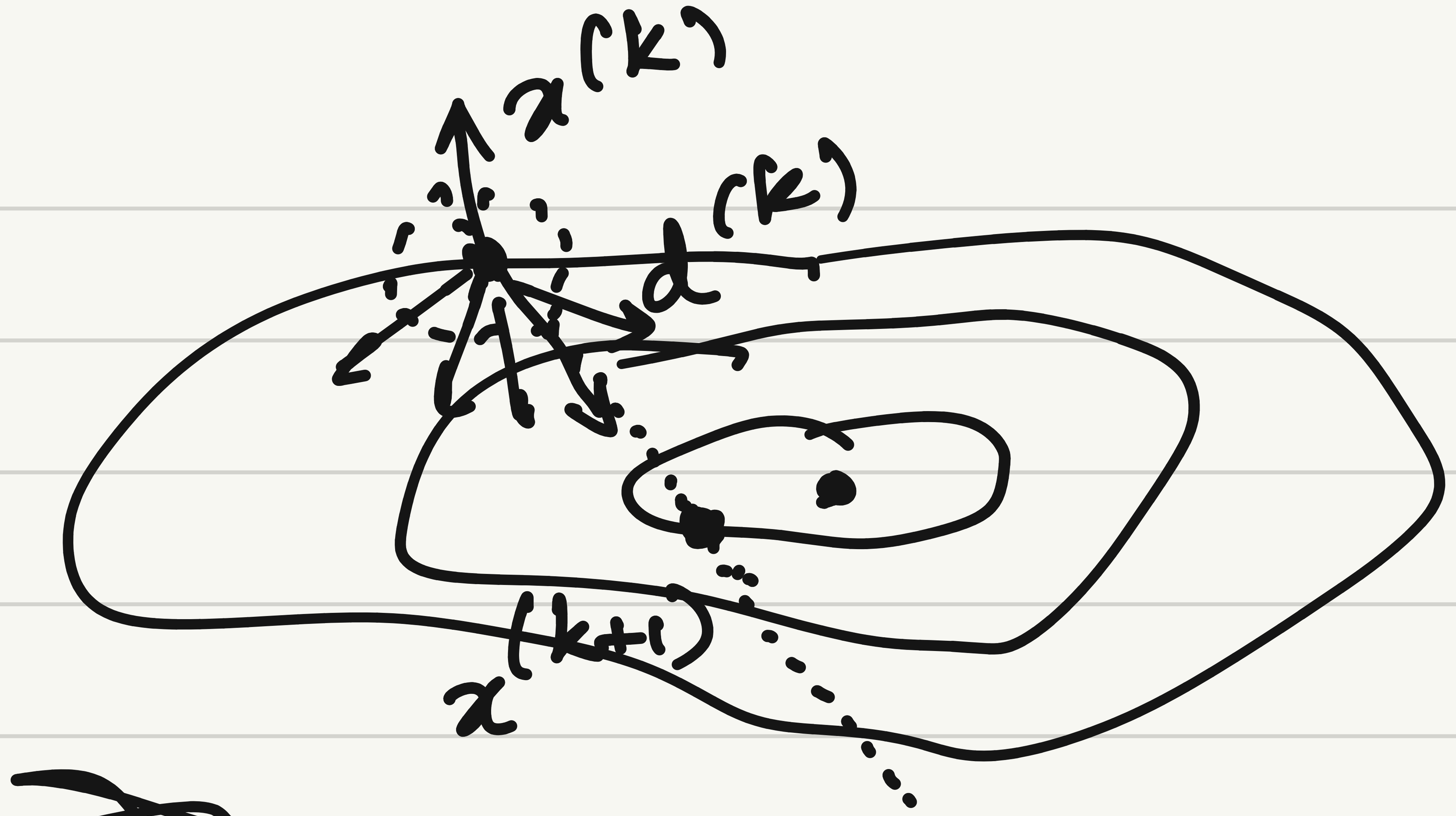
$f(x^{(k+1)}) < f(x^{(k)})$: $\boxed{\text{descent}}$

$x^{(0)}, x^{(1)}, x^{(2)}, \ldots \in \{x : f(x) \le f(x^{(0)})\} = \underline{S_0}$





$\mathbb{R}^2 \to \mathbb{R}$



$\boxed{\text{Descent method}}$

or $\boxed{\text{line search method}}$

At $x^{(k)}$, choose $\boxed{\text{search direction}}$ $d^{(k)}$,

choose $\boxed{\text{step size}}$ $t^{(k)}$ s.t. $f(x+td) \le f(x)$

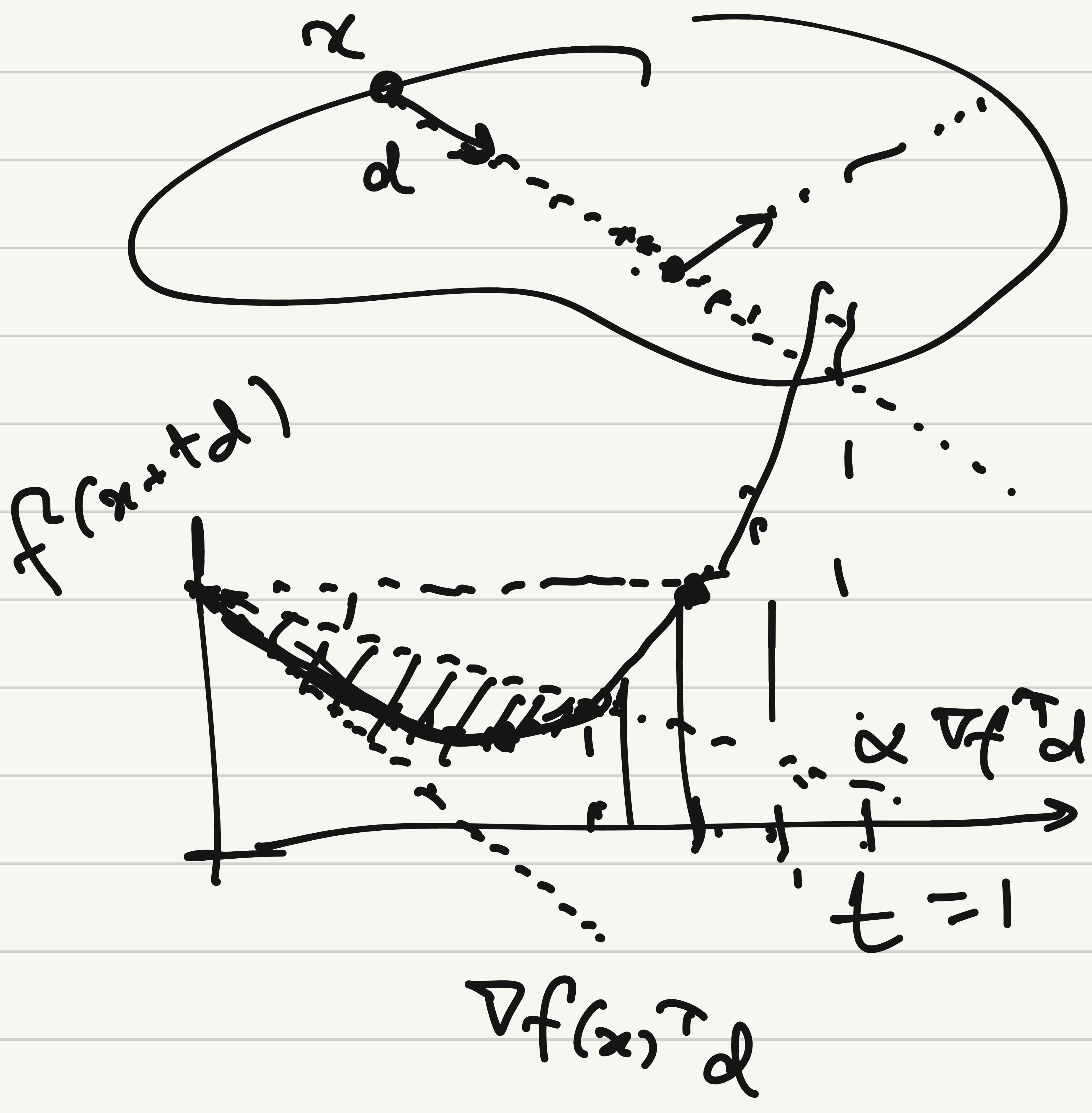
set $x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)}$

$\boxed{x^+ = x + td}$

$$\underbrace{f(x^+)}_{< f(x)} \approx f(x) + \underbrace{\nabla f(x)^T td}_{< 0}$$

$\boxed{\nabla f(x)^T d < 0}$ $\boxed{\text{descent direction}}$

$x,\ d,\ t = ?$



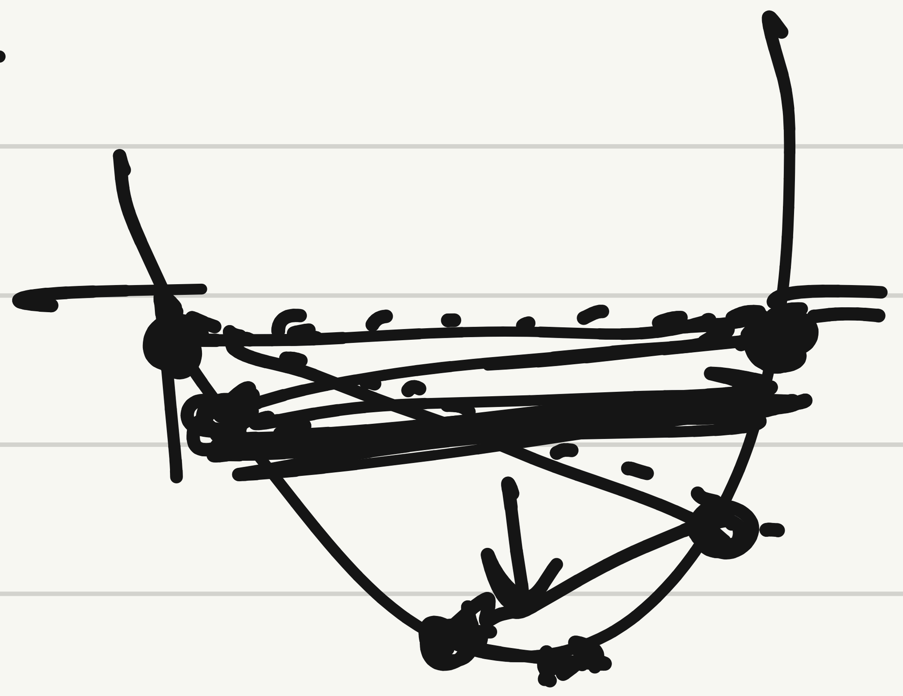
$\boxed{\text{Exact line search}}$: find exact minimizer of $f(x+td)$ s.t. $t \geq 0$

$\boxed{\text{Backtracking line search}}$:

start with large initial t

while $f(x+td) < f(x) + \alpha \nabla f(x)^T t d$:

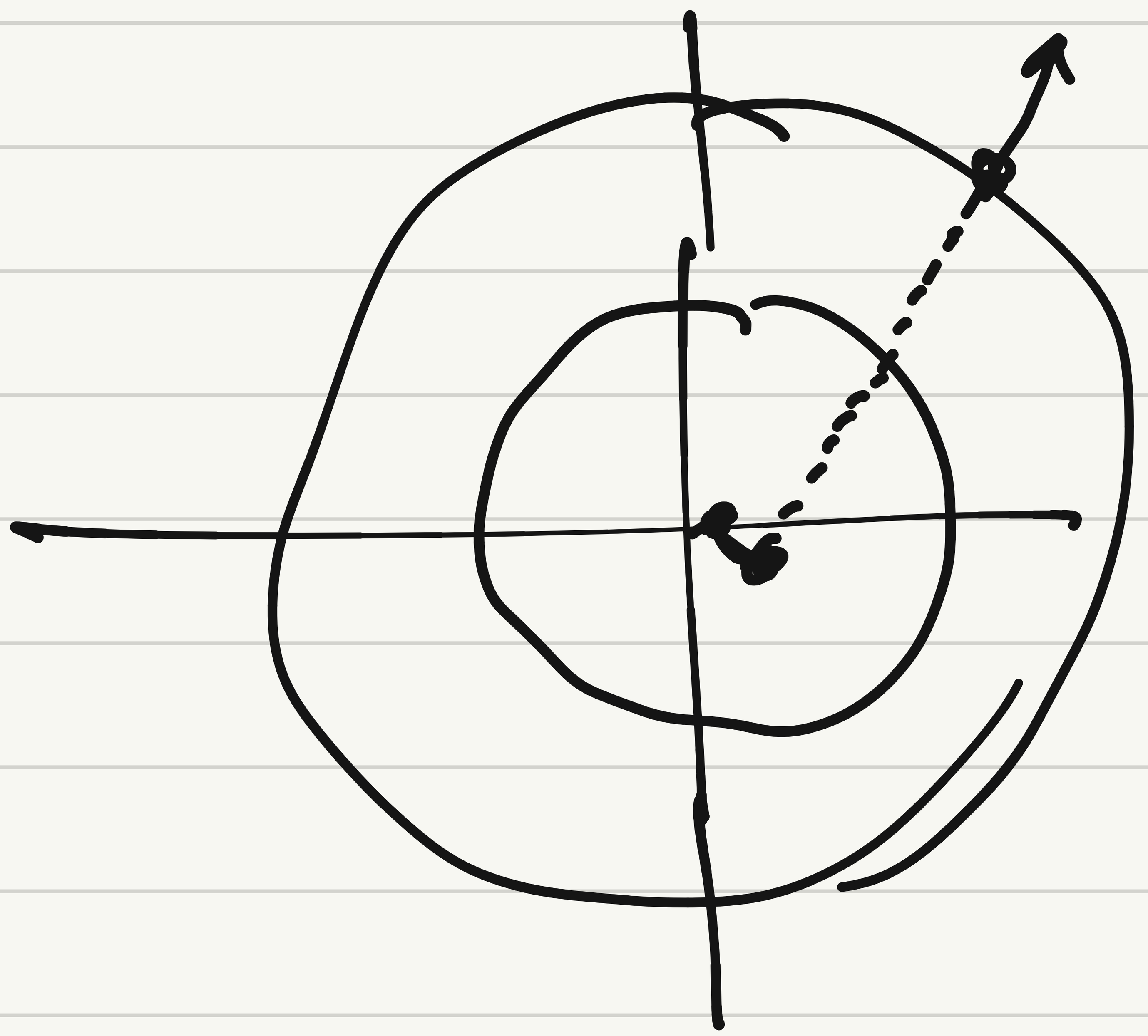
$t \leftarrow \beta t$



$0 < \beta < 1$ $\quad 0.1 \sim 0.8$

$0 < \alpha < 0.5$ $\quad 0.01 \sim 0.3$

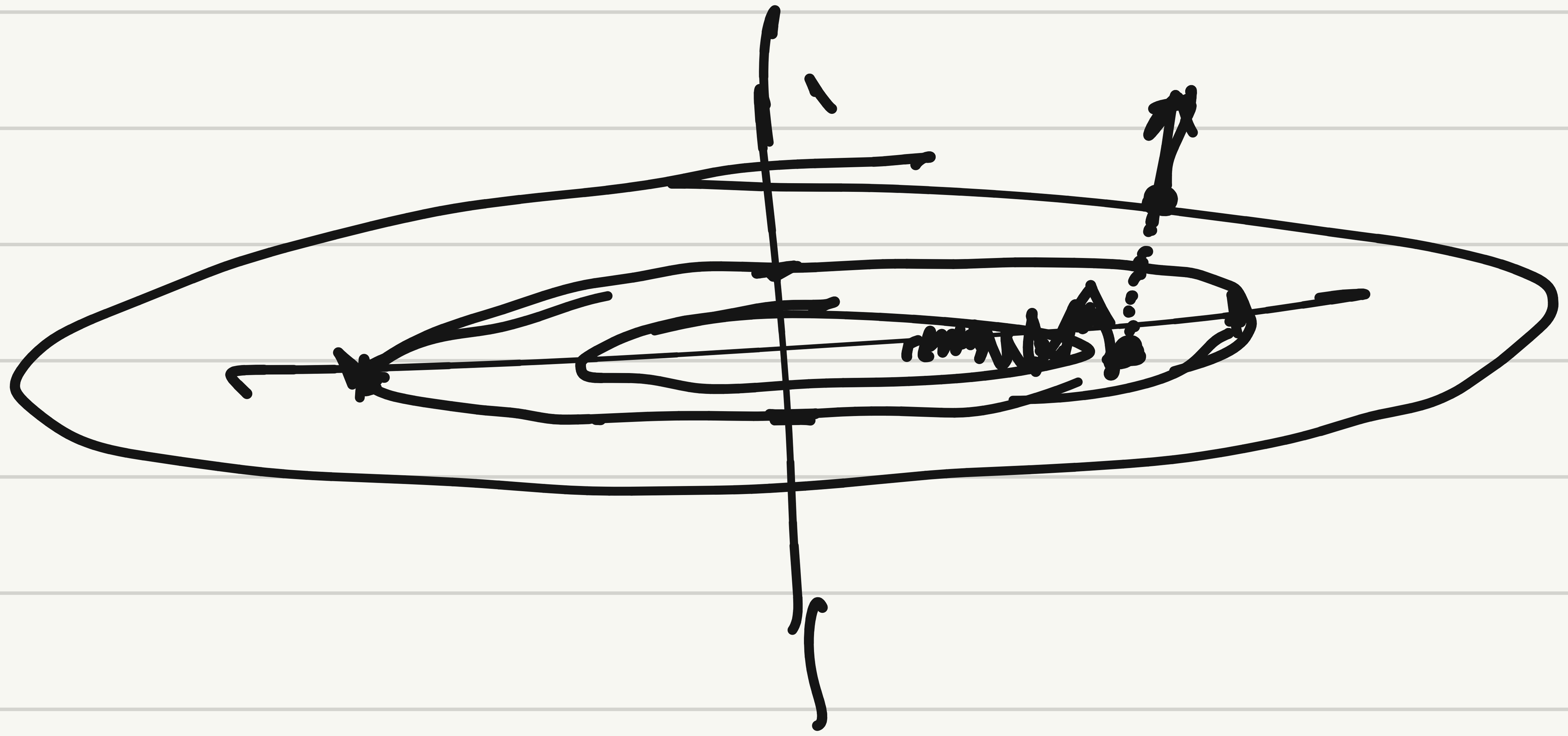
Choosing descent dir. d s.t. $\nabla f(x)^T d < 0$.



$d = -\nabla f(x)$: $\boxed{\text{gradient descent}}$

$$f(x) = \frac{1}{2} x^T \begin{bmatrix} 1 & \\ & \gamma \end{bmatrix} x = \frac{1}{2}(x_1^2 + \gamma x_2^2), \quad \gamma \geq 1$$

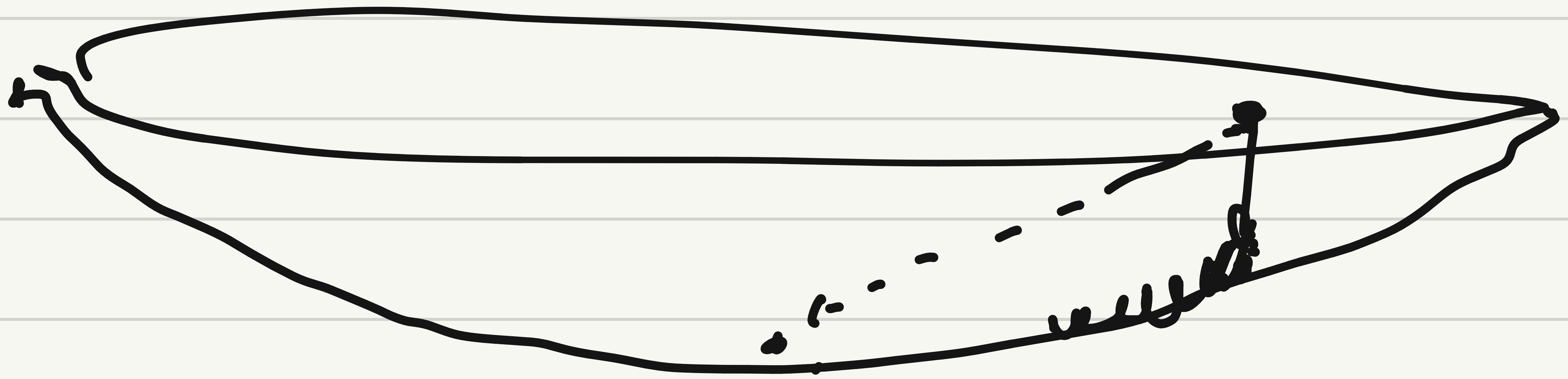
If $\gamma \gg 1$, $\nabla^2 f(x)$ badly conditioned,



Steepest descent

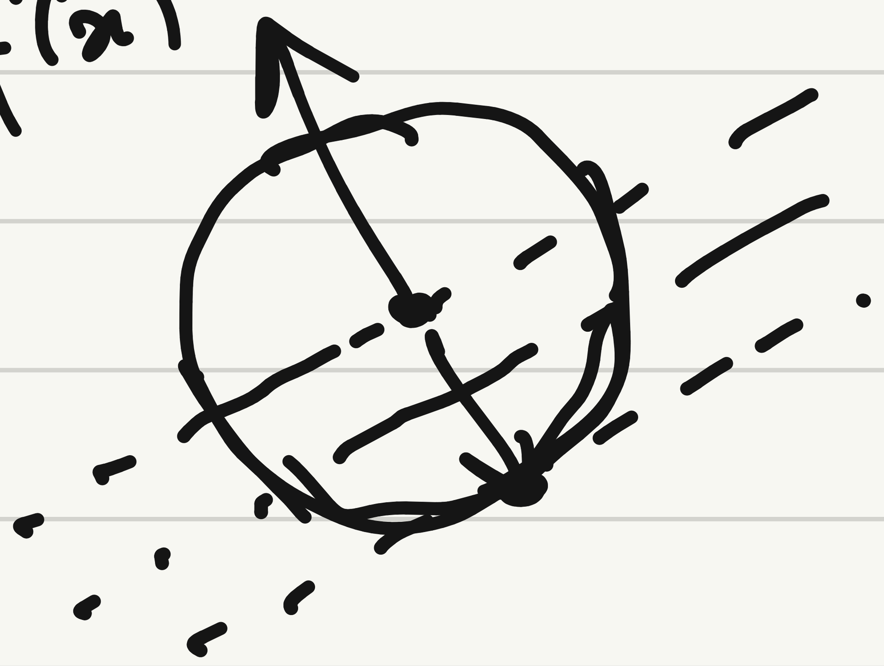
Choose d which descends fastest

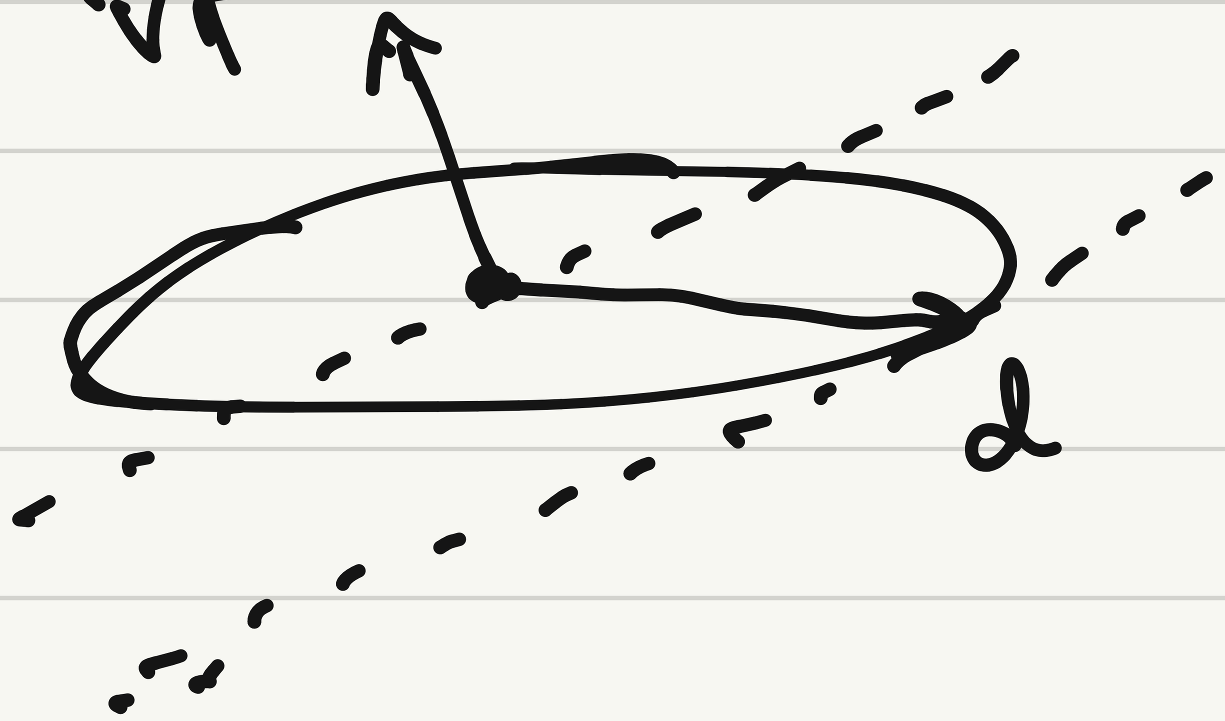
$$\min \nabla f(x)^T d \quad \text{s.t.} \quad \|d\| = 1$$



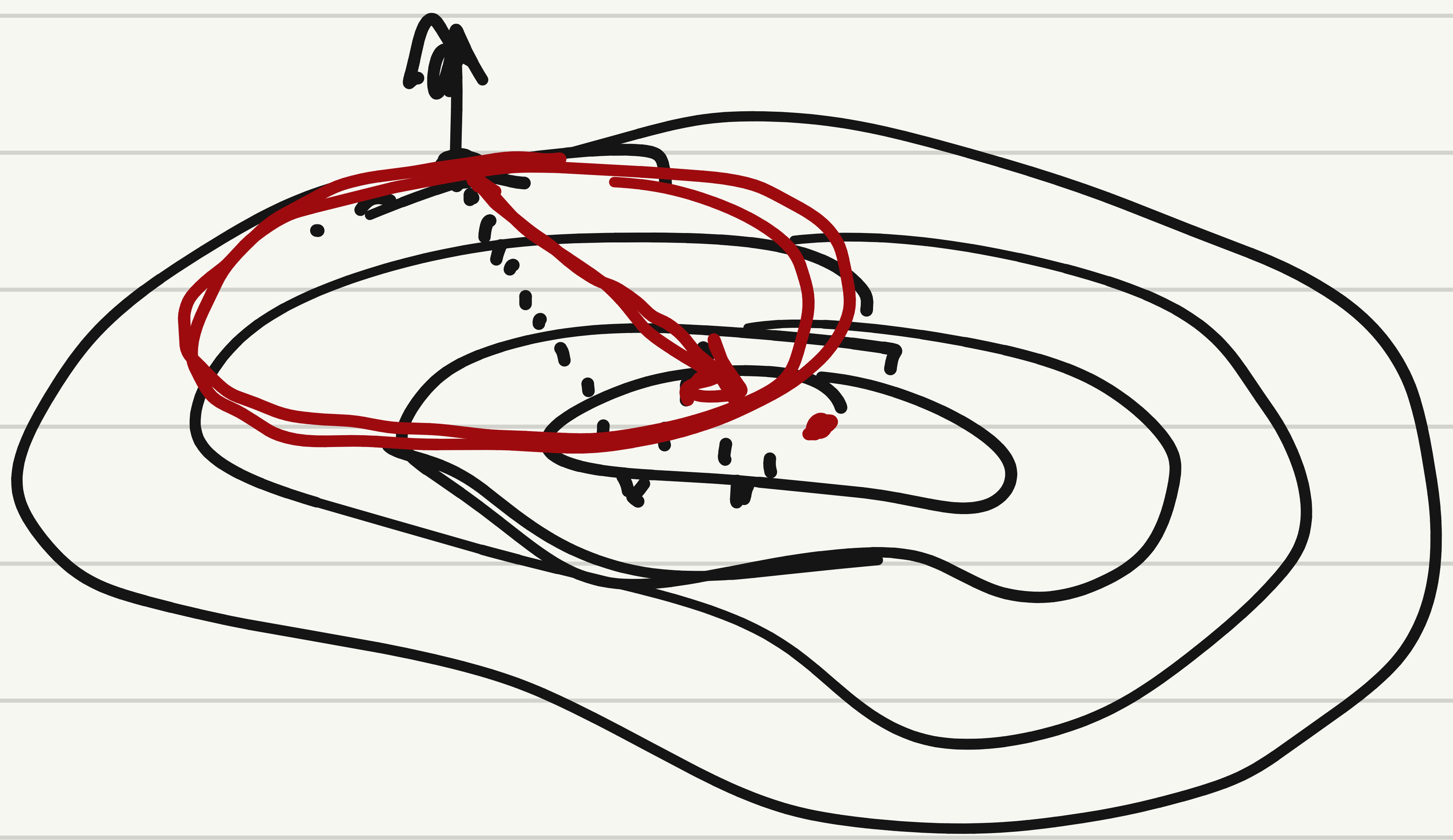
If norm is $\|\cdot\|_2 \implies$ gradient descent.

If $P \succ 0$, define $\|z\|_P = \sqrt{z^T P z} = \underbrace{\|Cz\|_2}$, where $\underline{C^T C} = P$





$$d = -\frac{P^{-1}\nabla f(x)}{\underbrace{\sqrt{\nabla f(x)^T P^{-1} \nabla f(x)}}} \quad (\text{normalized})$$

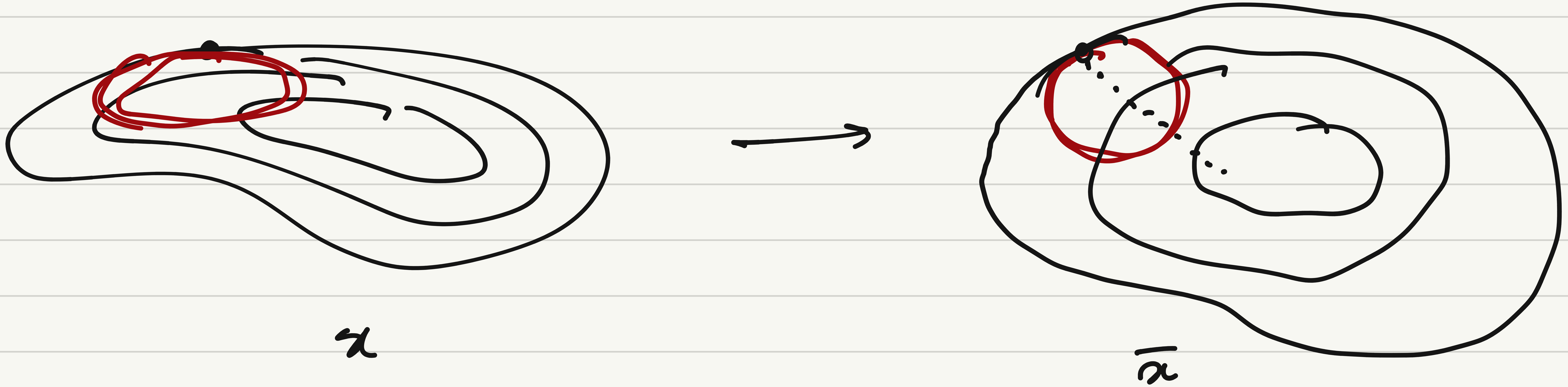


$$d = -P^{-1}\nabla f(x) \quad (\text{unnormalized})$$

Change of coords: $f(x) \longrightarrow \bar{x} = Cx$

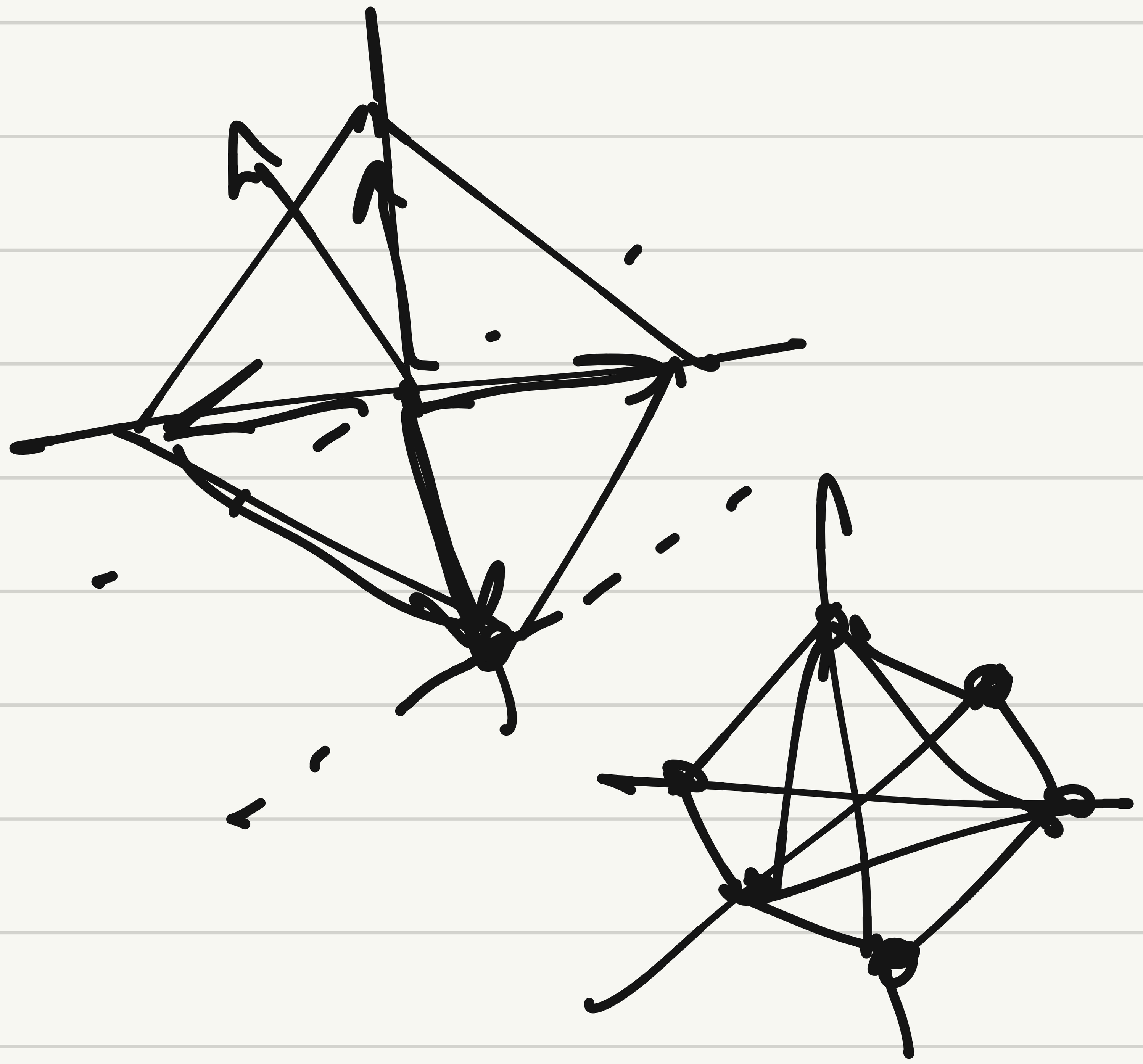
$$\bar{f}(\bar{x}) = f(x) = f(C^{-1}\bar{x})$$

$\|x\|_P = \|Cx\|_2 = \|\bar{x}\|_2 \longrightarrow$ steepest descent on x = gradient desc. on $\bar{x}$



$\|\cdot\|_p$

Choose d to min $\nabla f(x)^T d$ s.t. $\|d\|_1 = 1$



$d = \pm e_i$ i is the index with largest $(\nabla f(x))_i$

x^+ differs from x in only ith component!

coordinate descent

$\boxed{\text{Convergence analysis}}$

Assume f convex, twice diff., bounds on $\nabla^2 f(x)$

$x^{(0)}, x^{(1)}, \ldots \in S_0 = \{x : f(x) \leq f(x^{(0)})\}$

$\nabla^2 f(x) - mI \succeq 0$

$$\boxed{mI \preceq \nabla^2 f(x) \preceq MI} \quad \text{for all } x \in S_0$$

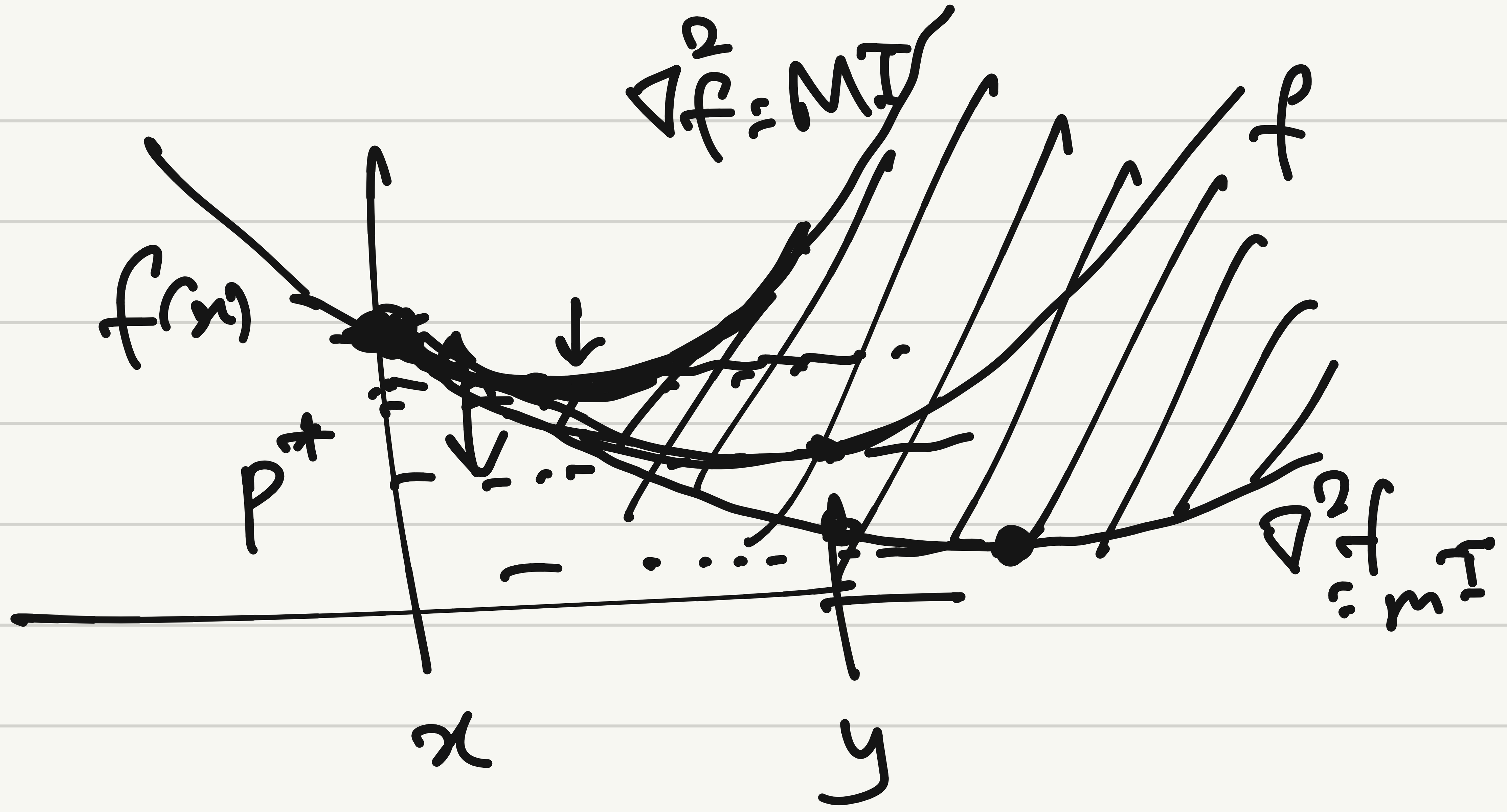
$\lambda_i \geq m \ \forall i$

$m > 0$

$\in \mathbb{R}$

f is $\boxed{\text{strongly convex}}$ (w/ convexity const. m)

$$mI \preceq \nabla^2 f(x) \preceq MI$$



$$f(x) + \nabla f(x)^T (y-x) + \frac{m}{2}\|y-x\|^2 \leftarrow$$
$$\leq f(y) :$$
$$\leq f(x) + \nabla f(x)^T (y-x) + \frac{M}{2}\|y-x\|^2$$

optimal value $p^* \geq f(x) - \frac{1}{2m}\|\nabla f(x)\|^2 \Rightarrow f(x) - p^* \leq \frac{1}{2m}\|\nabla f(x)\|^2$

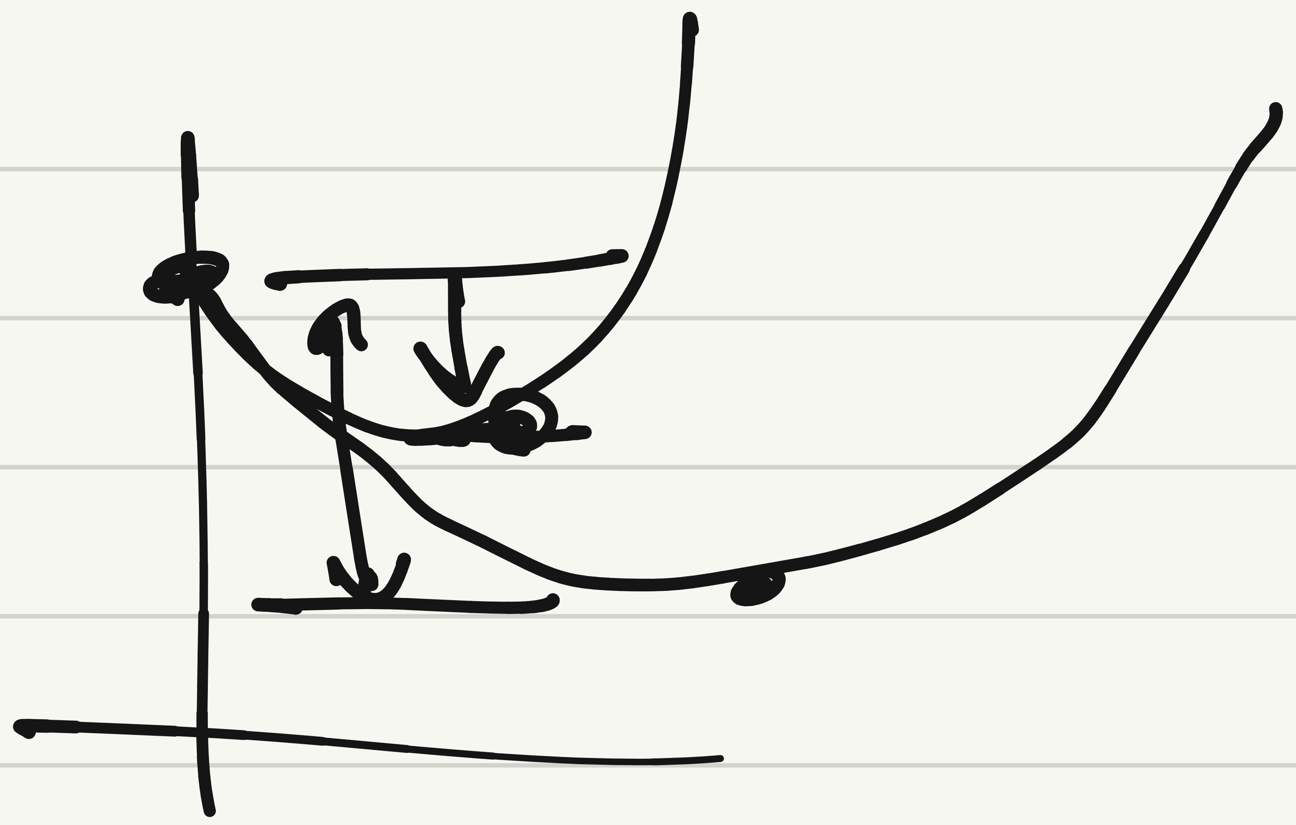
$\boxed{\text{Suboptimality}}$

for gradient descent, (w/ exact line search)

$$f(x^+) \leq f(x) - \underbrace{\frac{1}{2M}\|\nabla f(x)\|^2}_{}$$

grad. desc. reduces subopt by at least this subopt $\leq \frac{1}{2m} \|\nabla f\|^2$

progress of GD $\geq \frac{1}{2M} \|\nabla f\|^2$



$\Rightarrow$ each GD iter reduces subopt by factor of m/M

Convergence rate is linear with rate const. $\underline{(1 - m/M)}$

$\underline{\kappa(\nabla^2 f(x))} \leq M/m$

Backtracking: progress of GD $\geq \min\left(\alpha, \frac{\beta\alpha}{M}\right) \cdot \|\nabla f\|^2$