6. Numerical Algorithms

Let $f : \Omega \subseteq \mathbb{R}^d \to \mathbb{R}$ we study $\inf_{x \in A} f(x)$, $A \subseteq \Omega$

Given $\varepsilon > 0$, x^* solution; we look for $x^\varepsilon \in A$ s.t.

1. $|x^\varepsilon - x^*| \leq \varepsilon$ (error on solution)
2. $f(x^\varepsilon) \leq f(x^*) + \varepsilon$ (error on objective)

Strong convexity Suppose Ω convex. Let $\alpha \geq 0$

Def f is α-convex iff for all $x, y \in \Omega$, $t \in [0,1]$

$$f(x(1-t) + yt) \leq f(x)(1-t) + t f(y) - \alpha \frac{t(1-t)}{2} |x-y|^2$$

NB. f α-convex $\Rightarrow$ strict. convex $\Rightarrow$ convex

Remarks If f is C^1

f convex $\iff f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle \quad \forall x, y \in \Omega$

$\iff \langle \nabla f(x) - \nabla f(y), x-y \rangle \quad \forall x, y \in \Omega$

f str. convex $\iff f(y) > f(x) + \langle \nabla f(x), y-x \rangle \quad \forall x \neq y$

$\iff \langle \nabla f(x) - \nabla f(y), x-y \rangle > 0 \quad \forall x \neq y$

f α-convex $\iff f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle + \alpha \frac{|y-x|^2}{2}$

$\iff \langle \nabla f(x) - \nabla f(y), x-y \rangle \geq \alpha |x-y|^2$

Si $f \in C^2$, f α-convexe $\Leftrightarrow D^2 f(x) \geq \alpha I_d$

<u>Lemma</u> Let $f \in C^1(\mathbb{R}^d, \mathbb{R})$ α-convex and x^* be its unique minimizer then

$$\| x - x^* \| \leq \frac{2 \| \nabla f(x) \|}{\alpha}$$

$$f(x) - f(x^*) \leq \frac{\| \nabla f(x) \|^2}{2\alpha}$$

<u>NB</u>. the error in the optimality conditions controls the error on the minimum and the error in the minimizer.

$$f(x) \geq f(x^*) \geq f(x) + \langle \nabla f(x), x^* - x \rangle + \frac{\| x - x^* \|^2}{2} \alpha$$

We get the first by triangular inequality.

$$f(x) - f(x^*) + \alpha \frac{\| x - x^* \|^2}{2} \leq \| \nabla f(x) \| \, \| x^* - x \|$$

Cauchy-Schwarz. $\square$

Descent methods

Given $x^0 \in \Omega$ define

$$x^{n+1} = x^n + \tau^n d^n$$

where $\tau^n > 0$: step, $d^n \in \mathbb{R}^d$ is a descent direction x^n

Def $f \in C^0(\Omega; \mathbb{R})$, $d \in \mathbb{R}^d$ is a descent direct. at $x \in \Omega$ iff $\exists \bar{\tau}$: $\forall \tau \in (0, \bar{\tau})$ $f(x + \tau d) < f(x)$.

If f is diff. on Ω and $\langle d, \nabla f(x) \rangle < 0$ then d is a descent direction : Define

$$f(x+sd) = f(x) + s \langle \nabla f(x), d \rangle + o(s)$$

$\exists \bar{\tau}$: $\forall 0 \leq s \leq \bar{\tau}$, $o(s) < -s \langle \nabla f(x), d \rangle$ so

$$f(x+sd) < f(x) \qquad \forall s \in (0, \bar{\tau}).$$

Gradient descent with constant step-size

Descent direction $d = -\nabla f(x)$

Constant step size $\tau^n = \tau > 0$. $\forall n$.

Suppose: 1. $f \in C^1(\mathbb{R}^d)$

2. ∇f L-lip.

3. f bounded from below

$$f(x^{n+1}) = f(x^n) + \langle \nabla f(x^n), x^{n+1} - x^n \rangle$$

$$-\nabla f(x^n)$$

$$+ \int_0^1 \langle \nabla f(x^n(1-t) + x^{n+1}t), x^{n+1} - x^n \rangle \, dt$$

$$\leq f(x^{n}) - \tau \|\nabla f(x^n)\|^2 + L \frac{\|x^{n+1} - x^n\|^2}{2}$$

$$= f(x^n) - \tau\left(1 - \frac{L\tau}{2}\right) \|\nabla f(x^{n+1})\|^2$$

If $\tau < \frac{2}{L}$

$$\tau\left(1 - \frac{L\tau}{2}\right) \|\nabla f(x^n)\|^2 \leq f(x^n) - f(x^{n+1})$$

$$\tau\left(1 - \frac{L\tau}{2}\right) \sum_{n=0}^{N} \|\nabla f(x^n)\|^2 \leq f(x^0) - f(x^{N+1})$$

$$\leq f(x^0) - \inf_x f(x)$$

Letting $N \to \infty$ we deduce that

<u>Thm</u>. Let $f \in C^1(\mathbb{R}^d)$, ∇f L-lip, f bounded from below. Then if $\tau < \frac{2}{L}$

1. $f(x^n)$ is strictly decreasing
2. $\nabla f(x^n) \to 0$, $n \to \infty$.

Suppose that <u>additionally</u> f is convex:

$$f(x^{n+1}) \leq f(x^n) - \tau\left(1 - \frac{\tau L}{2}\right)\|\nabla f(x^n)\|^2$$

$$\leq f(x^*) - \langle \nabla f(x^n), x^* - x^n \rangle - \tau\left(1 - \frac{\tau L}{2}\right)\|\nabla f(x^n)\|^2$$

$\tau \leq \frac{1}{L}$

$$\leq f(x^*) - \langle \nabla f(x^n), x^* - x^n \rangle - \frac{\tau}{2}\|\nabla f(x^n)\|^2$$

$$= f(x^*) - \frac{1}{2\tau}\Big[\| x^* - x^n + \nabla f(x^n)\tau\|^2 - \|x^* - x^n\|^2\Big]$$

$$= f(x^*) - \frac{1}{2\tau}\Big[\|x^{n+1} - x^*\|^2 - \|x^n - x^*\|^2\Big] \quad (*)$$

$$\sum_{n=0}^{N} f(x^{n+1}) = (N+1) f(x^*) + \frac{1}{2\tau}\Big[\|x^0 - x^*\|^2 - \|x^{N+1} - x^*\|^2\Big]$$

We obtain since $f(x^n)$ is decreasing:

$$f(x^{N+1}) - f(x^*) \leq \frac{1}{2\tau(N+1)}\|x^0 - x^*\|^2$$

If f is α-convex, from $(*)$ we get

$$\alpha\frac{\|x^{n+1} - x^*\|^2}{2} \leq \frac{1}{2\tau}\|x^n - x^*\|^2 - \frac{1}{2\tau}\|x^{n+1} - x^*\|^2$$

$$\Rightarrow \|x^{n+1} - x^*\|^2 \leq \tau/(\tau\alpha + 1)\,\|x^n - x^*\|^2$$

thm Let $f \in C^1(\mathbb{R}^d)$ convex and suppose ∇f is L-lip. Let x^* be a minimizer of f. If $\boxed{\tau < \frac{2}{L}}$ $\exists\, C > 0$ s.t.

$$f(x^k) - f(x^*) \leq \frac{C}{k}$$

If f is α-convex $x^n \to x^*$ linearly
($\exists\, \lambda < 1$. s.t. $\|x^n - x^*\| \leq C\lambda^n$)

NB. If ∇f is L-lip and f is α convex : $\forall x, y$:

$$\alpha \|x-y\|^2 \leq \langle \nabla f(x) - \nabla f(y), x-y \rangle \leq L\|x-y\|^2$$

so $\boxed{\alpha \leq L}$

NB We can prove a similar result observing that

$$x^{n+1} = F(x^n) := x^n - \tau \nabla f(x^n)$$

and
$$\|F(x) - F(y)\|^2 = \|x-y\|^2 - 2\tau \langle x-y, \nabla f(x) - \nabla f(y) \rangle + \tau^2 \|\nabla f(x) - \nabla f(y)\|^2$$

$$\leq \underbrace{(1 - 2\tau\alpha + \tau^2 L^2)}_{<1} \|x-y\|^2$$

if $\tau < \frac{2\alpha}{L^2} < \frac{2}{L}$

$$\left.\begin{aligned} &\|x^{m+1}-x^*\| \le \lambda \|x^m - x^*\| \\ &\text{with } \lambda = \sqrt{1-2\tau\alpha+\tau^2L^2} \end{aligned}\right\} \text{ linear convergence}$$

<u>Exercise</u> Study case of quadratic functional

$$f(x) = \frac{x^T A x}{2} - b^T x + c \quad , A > 0$$

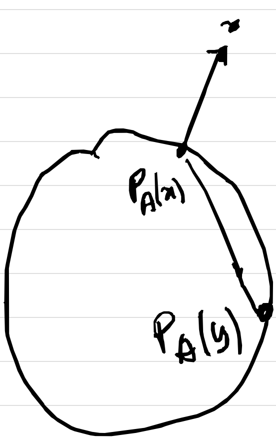
express constants α, L, λ in terms of the eigenvalues.

<u>Exercise</u> (Projected gradient method) Suppose A convex and closed and define

$$x^{m+1} = P_A\left(x^m - \tau \nabla f(x^m)\right)$$

$P_A(x) = \underset{y\in A}{\operatorname{argmin}} \|x-y\|^2$ is the orthog. projection.

$\|P_A(x) - P_A(y)\| \le \|x-y\|$



$$\langle x - P_A(x), P_A(y) - P_A(x) \rangle \le 0$$

$$\langle y - P_A(y), P_A(x) - P_A(y) \rangle \le 0$$

$$\langle x - y - P_A(x) + P_A(y), P_A(y) - P_A(x) \rangle \le 0$$

$$\|P_A(y) - P_A(x)\| \le \|x-y\|$$

Backtracking and optimal step size

Exact line search:

$$x^{n+1} \in \underset{\tau}{\operatorname{argmin}}\ f(x^n + \tau d^n)$$

Descent condition

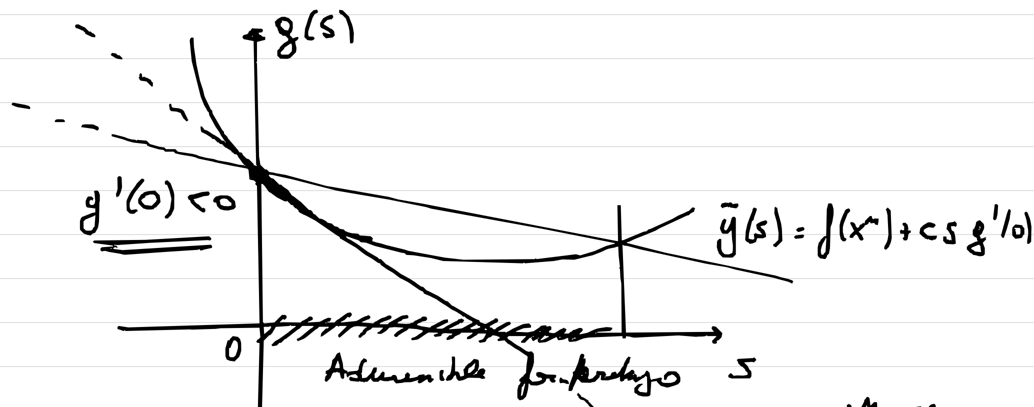
Can we do better?

while $f(x^{n+1}) \geq f(x^n)$:

$\tau^n \leftarrow \beta \tau^n$ $\qquad \beta < 1$

$x^{n+1} \leftarrow x^n + \tau^n d^n$

Armijo condition let $g(s) := f(x^n + s d^n)$



$$y(s) = f(x^n) + s \langle d^n, \nabla f(x^n) \rangle$$

If f is convex $g \geq y$. However for $c < 0$

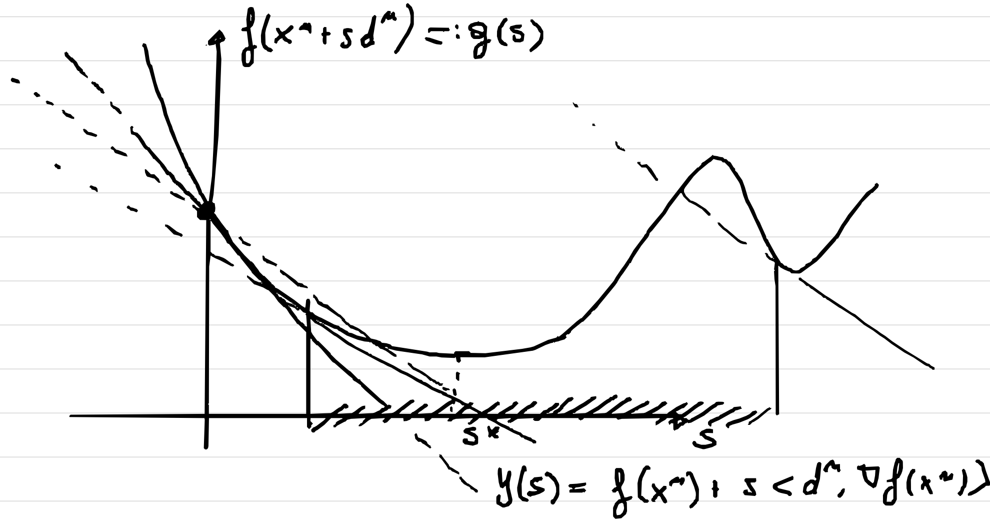
$\exists \bar{s} : \forall s \in (0, \bar{s}) \quad g(s) < \tilde{y}(s)$

So we can require

$$f(x^{n+1}) \leq f(x^n) + c_A \underbrace{\langle d^n, \nabla f(x^n) \rangle}_{<0}$$

We are using the information that d^n is a descent direction to come up with a better criterion.

Wolfe-conditions Avoid small step-sizes



At the optimum $g'(s^*) = 0$ and $g'(0) < 0$

To avoid small step-sizes require

$$g'(s) > c_W \, g'(0) \qquad \text{where } c_W < 1$$

$$\boxed{\langle \nabla f(x^n + s d^n), d^n \rangle > c_W \langle \nabla f(x^n), d^n \rangle}$$

<u>Newton's method</u>. $f : \mathbb{R}^d \to \mathbb{R}$ smooth

$$f(x) \approx f(x^n) + \langle \nabla f(x^n), x - x^n \rangle + \frac{1}{2} \langle D^2 f(x^n)(x-x^n), x-x^n \rangle$$

Define x^{n+1} as minimizer of the quadratic function

$$x \mapsto \frac{\langle Ax, x \rangle}{2} + \langle b, x \rangle + c$$

$$A = D^2 f(x^n), \quad b = \nabla f(x^n)$$

Suppose $A \succ 0$ (positive definite) then problem admit a unique minimizer (strongly convex, smooth function)

Optimality conditions $\quad x = -A^{-1} b$

$$\boxed{\begin{aligned} d^n &= -[D^2 f(x^n)]^{-1} \nabla f(x^n) \\ x^{n+1} &= x^n + d^n \end{aligned}}$$

<u>NB</u>. Only well-defined as long as $D^2 f(x^n)$ invertible.

Assumptions: let x^* a local minimizer.

1. f is $\mathcal{C}^2(\mathbb{R}^d)$
2. $D^2 f$ is γ-Lipschitz
3. $D^2 f(x^*) \succeq \alpha \, Id$ (locally strongly convex)

For the last point observe that

$$\|D^2 f(x) - D^2 f(x^*)\| \leq \gamma \|x - x^*\|$$

induced norm (largest singular value)

$$\Rightarrow \quad -\gamma \|x - x^*\| \, Id \leq D^2 f(x) - D^2 f(x^*)$$

$$\Rightarrow \quad D^2 f(x) \geq (\alpha - \gamma \|x - x^*\|) \, Id$$

$$\|D^2 f(x)^{-1}\| \leq (\alpha - \gamma \|x - x^*\|)^{-1}, \quad \|x - x^*\| < \frac{\alpha}{\gamma}$$

If x^n is in this ball $D^2 f(x^n)$ is <u>invertible</u> and

$$x^{n+1} = x^n - D^2 f(x^n)^{-1} \nabla f(x^n) \quad \text{well-defined}$$

$$x^{n+1} - x^* = x^n - x^* - D^2 f(x^n)^{-1} \int_0^1 D^2 f(x^* + s(x^n - x^*))(x^n - x^*)$$

$$= D^2 f(x^n)^{-1} \int_0^1 \left[D^2 f(x^n) - D^2 f(x^* + s(x^n - x^*)) \right] (x^n - x^*)$$

$$\|x^{n+1} - x^*\| \leq \frac{1}{(\alpha - \gamma \|x^n - x^*\|)} \frac{1}{2} \gamma \|x^n - x^*\|^2$$

$$\|x^n - x^*\| \leq \frac{2\alpha}{3\gamma} \quad \Rightarrow \quad \|x^{n+1} - x^*\| \leq \frac{2\alpha}{3\gamma}$$

All iterates in the same ball

Thm Let $f \in C^2(\mathbb{R}^d)$ s.t. $D^2 f$ is γ-Lip, let x^* a local minimizer with $D^2 f(x^*) \succeq \alpha \, Id$, $\alpha > 0$. then if $\|x^0 - x^*\| \leq \frac{2\alpha}{3\gamma}$ the iterates are well defined and converge quadratically towards x^*.

Remarks

1. Only local convergence!
2. At each iteration we need to solve a linear system.