2. Differentiable optimization pbs

V normed vect. space $\quad$ neighborhood

$U \subset V$ convex $\quad F : \mathcal{O}(U) \to \mathbb{R}$ differentiable at $\hat{u} \in U$

If F has a min at $\hat{u}$ relat. to U

then $\langle \nabla F(\hat{u}), u - \hat{u} \rangle \geq \; \forall u \in U$ $\quad$ ⊛

If in addition F is convex $(\Longleftarrow)$

If U open $\quad$ ⊛ $\quad \nabla F(\hat{u}) = 0$

1. Relaxation methods $\quad F : \mathbb{R}^n \to \mathbb{R}$

minimize with respect to one component at a time

$\forall k$ (iteration number) $\quad \forall i$ (step $1, \dots, n$)

$$F(u_{k+1}(1), \dots, \underline{u_{k+1}(i)}, u_k(i+1) \dots u_k(n))$$

$$= \inf_{t \in \mathbb{R}} F(u_{k+1}(1), \dots, u_{k+1}(i-1), \underline{t}, u_k(i+1), \dots, u_k(n))$$

Thm. $\underline{F \sim C^1}$ strictly convex, coercive

$\Rightarrow (u_k)_{k \geq 0}$ converges to the minimizer

the differentiability of F – key role

counter-example: $\quad F(u_1, u_2) = u_1^2 + u_2^2 - 2(u_1 + u_2) + 2|u_1 - u_2$

$F(-1, -1) = -2$.

Apply the method starting with $(0,0)$

$F(t, 0) = t^2 - 2t + 2|t|$ hence $u_1 = 0$ and so on....

the sequence in 0 all the time!

for $F(u) = \frac{1}{2}\langle Bu, u \rangle - \langle c, u \rangle$ Rel = Gauss-Seidel.

F ellip. notice $u_{k+1} = u_k - \overbrace{\rho(u_k)}^{\neq 0} \nabla F(u_k)$ $\qquad \nabla F(u_k) = \dfrac{u_k - u_{k+1}}{\rho}$

$\langle \nabla F(u_{k+1}), u_k - u_{k+1} \rangle = 0$

F ellip. $\Longrightarrow$ $\underline{F(u_k) - F(u_{k+1})} \geqslant \overbrace{\langle \nabla F(u_{k+1}), u_k - u_{k+1} \rangle}^{0}$
$$+ \underline{\frac{\mu}{2} \| u_k - u_{k+1} \|^2}$$

2) $F(u_k)$ decreasing lower bounded (by $F(\hat{u})$)

$\Rightarrow F(u_k)$ converges $\lim\limits_{k \to \infty} \underline{\left(F(u_k) - F(u_{k+1}) \right)} = 0$

$\Longrightarrow \lim\limits_{k \to \infty} \| u_{k+1} - u_k \| = 0$ $\quad *$

Remind: F elliptic $\Rightarrow$ F strictly convex coercive and
$F(u) - F(v) \geqslant \langle \nabla F(v), u - v \rangle + \frac{\mu}{2} \| u - v \|^2$

3) $\| \nabla F(u_k) \|^2 = \langle \nabla F(u_k), \nabla F(u_k) \rangle - \overline{\langle \nabla F(u_k), \nabla F(u_{k+1}) \rangle}^{\,0}$
$= \langle \nabla F(u_k), \nabla F(u_k) - \nabla F(u_{k+1}) \rangle$
$\leq \| \nabla F(u_k) \| \, \| \nabla F(u_k) - \nabla F(u_{k+1}) \|$

$\| \nabla F(u_k) \| \leq \| \nabla F(u_k) - \nabla F(u_{k+1}) \|$ $\quad *$

4) $F(u_k)$ $\underline{\text{decreasing}}$, F coercive $\exists\, r > 0$ such that
$u_k \in \overline{B}(0, r)$ $\quad \forall k$ $\quad$ (u_k bounded)
$F \sim C^1 \Rightarrow \nabla F$ uniformly continuous on $\overline{B(0,r)}$ compact

$\forall \varepsilon$ $\exists N$ sth $k \geqslant N \Rightarrow$ we have
$* \quad \| u_k - u_{k+1} \| \leq \eta \Longrightarrow \| \nabla F(u_k) - \nabla F(u_{k+1}) \| < \varepsilon$

$\Longrightarrow \lim \| \nabla F(u_k) \| = 0$

5) $\mu \| u_k - \hat{u} \|^2 \overset{\text{ellip.}}{\leq} \langle \nabla F(u_k) - \overbrace{\nabla F(\hat{u})}^{0}, u_k - \hat{u} \rangle$
$= \langle \nabla F(u_k), u_k - \hat{u} \rangle \leq \| \nabla F(u_k) \| . \| u_k - \hat{u} \|$

$\| u_k - \hat{u} \| \leq \left(\frac{1}{\mu} \right) \| \nabla F(u_k) \| \longrightarrow 0$

2.2 Gradient based methods

make $F(u_k) - F(u_{k+1})$ as large as possible

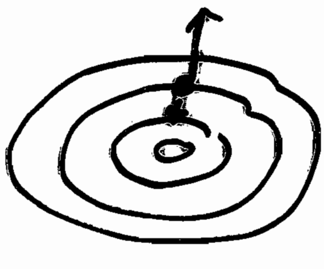
$$F(u_k + h) = F(u_k) + \langle \nabla F(u_k), h \rangle + \|h\| \underbrace{\varepsilon(h)}_{\xrightarrow[h\to 0]{} 0}$$

$$F(u_k) - F(u_k + h) =$$

$$= -\langle \nabla F(u_k), h \rangle - \|h\| \varepsilon(h)$$

$$\le \|\nabla F(u_k)\| \cdot \|h\| - \|h\| \varepsilon(h)$$

Schwarz inequality reached if $h = -\rho \nabla F(u_k)$, $\rho > 0$



the level lines

(a) Grad. with optimal step

$$\begin{cases} F\big(u_k - \underline{\rho(u_k)} \nabla F(u_k)\big) = \inf\limits_{\rho \in \mathbb{R}} F\big(u_k - \rho \nabla F(u_k)\big) \\ u_{k+1} = u_k - \rho(u_k) \nabla F(u_k) \end{cases}$$

Thm. $F: \mathbb{R}^n \to \mathbb{R}$ elliptic $\Rightarrow u_k \longrightarrow \hat{u}$ (the solution)

Proof (in several steps)

1) $\nabla F(u_k) \neq 0 \quad \forall k \ge 0$ (otherwise we have the sol.!)

define $f_k(\rho) = F(u_k - \rho \nabla F(u_k))$ $\quad \rho \in \mathbb{R}$

f_k - coercive, strictly convex, has one (ex.) minimizer $\rho(u_k)$

$f_k'(\rho) = 0$

$$f_k'(\rho) = -\langle \nabla F(u_k), \nabla F(u_k - \underbrace{\rho}_{\rho(u_k)} \nabla F(u_k)) \rangle = 0$$

$$\langle \nabla F(u_k), \nabla F(u_{k+1}) \rangle = 0 \quad \boxed{\text{consecutive directions } \perp}$$

Quadratic elliptic

$F(u) = \frac{1}{2} \langle Bu, u \rangle - \langle c, u \rangle$ B real sym positive def $c \in R^n$

$\nabla F(u) = Bu - c$

Find the direction and the opt. step.
By applying the method.

$f_k(\rho) = F(u_k - \rho \underbrace{(Bu_k - c)}_{\nabla F(u_k)})$, find the sol. of $f_k'(\rho) = 0$.

~~$d_k = \nabla$~~

Rmk: the thm. holds true if F is C^1 strictly convex and coercive.

Q linear convergence speed.

(B) Gradient with a variable step

$u_{k+1} = u_k - \underline{\rho_k} \nabla F(u_k)$

Thm. F differentiable
Suppose $\exists \mu > 0$ $M > 0$ such that

$\langle \nabla F(u) - \nabla F(v), u - v \rangle \geq \mu \|u - v\|^2 \quad \forall u, v$

$\|\nabla F(u) - \nabla F(v)\| \leq M \|u - v\| \quad \forall u, v$

Let $\xi \in (0, \mu/M^2)$ (a constant)

If $\boxed{\xi \leq \rho_k \leq \frac{2\mu}{M^2} - \xi}$ $\forall k \geq 0$ then (u_k) converges and sufficient

$\|u_k - \hat{u}\| \leq \gamma^k \|u_0 - \hat{u}\|$

$\gamma = \sqrt{\cdots} < 1$

3 Conjugate gradient method

$F(u) = \frac{1}{2}\langle Bu, u\rangle - \langle c, u\rangle$ $B = B^T$ positive def.
$u \in \mathbb{R}^n$

$\nabla F(u) = Bu - c$

$$\boxed{\begin{aligned} F(u_{k+1}) &= \inf_{u \in u_k + H_k} F(u) \\ H_k &= \operatorname{Span}\{\nabla F(u_i),\ 0 \le i \le k\} \end{aligned}}$$

$$H_k = \left\{ g(\alpha) = \sum_{i=0}^{k} \alpha_i^{k} \nabla F(u_i) : \alpha_i^k \in \mathbb{R} \quad 0 \le i \le k \right\}$$

$$F(u_{k+1}) = \inf_{\alpha \in \mathbb{R}^k} \underbrace{F(u_k + g(\alpha))}_{f(\alpha)} = F(u_k + g(\hat{\alpha}))$$

$\hat{\alpha}$ is the unique solution of $\nabla f(\alpha) = 0$

$$\frac{\partial f(\alpha)}{\partial \alpha_i} = 0 = \langle \nabla F(u_{k+1}), \nabla F(u_i)\rangle$$
$$u_{k+1} = u_k + g(\hat{\alpha})$$

$$\forall k \quad \langle \nabla F(u_{k+1}), \nabla F(u_i)\rangle = 0 \quad 0 \le i \le k$$

$$\Rightarrow \langle \nabla F(u_{k+1}), u\rangle = 0 \quad \forall u \in H_k$$

k $\{\nabla F(u_i)\}$ are linearly independent.

if $\nabla F(u_k) = 0$ then $\hat{u} = u_k$

if $\nabla F(u_k) \neq 0$ then $\dim H_k = k+1$

<u>Thm.</u> The method converges at n iterations at most.

Suppose $\nabla F(u_{n-1}) \neq 0$ then $H_{n-1} = \mathbb{R}^n$ spanned by n lin.-indep. vect.

$\langle \nabla F(u_n), u\rangle = 0 \quad \forall u \in \mathbb{R}^n$

$\Leftrightarrow \nabla F(u_n) = 0$

The iterates. $\nabla F(u_k) \neq 0$

$$\begin{cases} u_{k+1} = u_k - \rho\, d_k \\ d_k = \sum_{i=0}^{k} \alpha_i^k \nabla F(u_i) = \sum_{i=0}^{k-1} \alpha_i^k \nabla F(u_i) + \nabla F(u_k) \end{cases} \qquad \nabla F(u) = Bu - c$$

$$\underline{\nabla F(u_{k+1})} = \nabla F(u_k - \rho d_k)$$

$$= \underline{\nabla F(u_k) - \rho B d_k}$$

$$\langle \nabla F(u_{k+1}), \nabla F(u_k) \rangle = 0$$

$$= \| \nabla F(u_k) \|^2 - \rho_k \overbrace{\langle B d_k, \nabla F(u_k) \rangle}^{\neq 0}$$

$$0 \leq i \leq k-1$$

$$0 = \langle \nabla F(u_{k+1}), \nabla F(u_i) \rangle$$

$$= \underbrace{\langle \nabla F(u_k), \nabla F(u_i) \rangle}_{0} - \rho_k \langle B d_k, \nabla F(u_i) \rangle$$

$$\Rightarrow \quad \langle \underline{B d_k, \nabla F(u_i)} \rangle = 0$$

$$0 \leq j \leq k-1$$

$$\langle B d_k, d_j \rangle = \langle B d_k, \sum_{i=0}^{j} \alpha_i^j \nabla F(u_i) \rangle =$$

$$= \sum_{i=0}^{j} \alpha_i^j \langle \underline{B d_k, \nabla F(u_i)} \rangle = 0$$

$$\Rightarrow \quad \langle B d_k, d_j \rangle = 0 \qquad 0 \leq j \leq k-1$$

using also that B positive definite

$\Rightarrow \{ d_i \}$ are linearly independent

$$\underbrace{[\, d_0 \,\vdots\, d_1 \,\vdots\, \cdots \,\vdots\, d_k \,]}_{\text{rank} = k} = \underbrace{[\, \nabla F(u_0) \,\vdots\, \cdots \nabla F(u_k) \,]}_{\text{rank} = k} {}_{k}\!\begin{bmatrix} \alpha & & \\ & \ddots & \\ 0 & & \end{bmatrix}^{k}$$

all diag. elem. are $\neq 0$.

$0 \leq j \leq k-1$

$\langle B d_k, d_j \rangle = \langle B d_j, d_k \rangle = 0$

$= \langle \underbrace{\rho_j B d_j}_{\nabla F(u_j) - \nabla F(u_{j+1})}, d_k \rangle = \langle \nabla F(u_j) - \nabla F(u_{j+1}), d_k \rangle$

insert def of d_k.

$\alpha_j^k \cancel{\|\nabla F(u_j)\|}^2$

$= \sum_{i=0}^{k-1} \alpha_i^k \langle \nabla F(u_j), \nabla F(u_i) \rangle + \underbrace{\langle \nabla F(u_j), \nabla F(u_k) \rangle}_{0}$

$- \sum_{i=0}^{k-1} \alpha_i^k \langle \nabla F(u_{j+1}), \nabla F(u_i) \rangle -$

$- \langle \nabla F(u_{j+1}), \nabla F(u_k) \rangle \quad = 0$

$j = k-1 \qquad \alpha_{k-1}^k \|\nabla F(u_{k-1})\|^2 - \|\nabla F(u_k)\|^2 = 0$

$\alpha_{k-1}^k \|\nabla F(u_{k-1})\|^2 = \|\nabla F(u_k)\|^2$

$j = k-2 \quad \ldots$

$0 \leq j \leq k-2 \qquad \alpha_j^k \|\nabla F(u_j)\|^2 = \alpha_{j+1}^k \|\nabla F(u_{j+1})\|^2$

$= \|\nabla F(u_k)\|^2$

$$\alpha_j^k = \frac{\|\nabla F(u_k)\|^2}{\|\nabla F(u_j)\|^2} \qquad 0 \leq j \leq k-1.$$