3.6. The least squares problem.

$$f(x)=\sum_{k=1}^{n} r_k(x)^2 = \|r(x)\|^2, \quad r=\begin{bmatrix} r_1\\ r_2\\ \vdots \\ r_n\end{bmatrix}$$

Special, but often used problem.

Linear case: $r(x) = Ax - b$

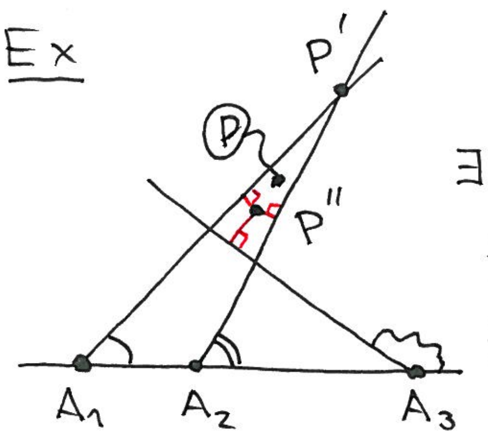
where $A = m\overset{n}{\square}$, $b = m\overset{1}{\square}$ – given

and $x \in \mathbb{R}^n$ is unknown.

exists

- $\exists x \in \mathbb{R}^n : Ax = b \Rightarrow \min_x f(x) = 0$
- $\nexists x \in \mathbb{R}^n \Rightarrow$ find $x \in \mathbb{R}^n : Ax \approx b$

Ex



2 measurements: $\exists x = P'$, Bad estimation

3 measurements: $\nexists x$, but $P'' \approx P$.

Solution: $f(x) = \|Ax - b\|^2 =$

$$= x^T A^T A x - 2 b^T A x + b^T b.$$

Quadratic function!

Stationary point: $\nabla f = 0$

$$\nabla f(x) = 2A^TAx - 2A^Tb = 0 \Leftrightarrow$$

$$\Leftrightarrow A^TAx = A^Tb \quad \text{– normal equation } (*)$$

Lemma: $\bar{x}$ solves $\min_x \|Ax - b\|^2 \Leftrightarrow$

$\Leftrightarrow \bar{x}$ is a solution to $(*)$.

Proof: $\boxed{\Rightarrow}$ see above.

$\boxed{\Leftarrow}$ $\bar{x}$ solves $(*)$ $\Rightarrow$ $\bar{x}$ is stationary point $\Rightarrow$

$$\overset{\text{Ex. 1.56}}{\Longrightarrow} f(x) = (x - \bar{x})^T A^T A (x - \bar{x}) + f(\bar{x}) =$$

$$= \|A(x - \bar{x})\|^2 + f(\bar{x}) \geq f(\bar{x}) \quad \text{and}$$

equality when $x = \bar{x} \Rightarrow \bar{x}$ optimal ∎

Remark: if A^TA is invertible then

$$x = \underbrace{(A^TA)^{-1}A^T}_{A^+ \,:\, \text{pseudoinverse}} b \quad \text{- unique solution}$$

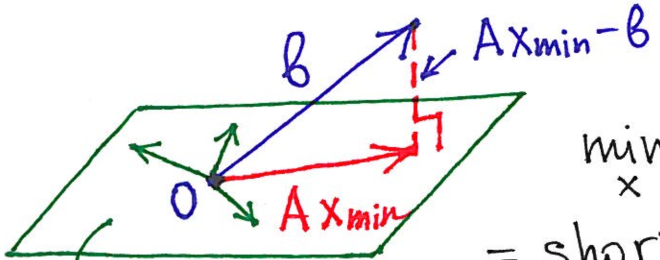
- if A^TA not invertible then

$$x = A^+ b \quad \text{- one of many solutions}$$

Geometrical interpretation of LS:

$$Ax = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{k=1}^{n} x_k \vec{a}_k$$

↑ linear combination of columns of A



$Lin(\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n)$

$\min\limits_x \|b - Ax\|_2 =$ = shortest distance => orthogonal =>

$$\Rightarrow \boxed{Ax_{min} - b \perp Ax, \ \forall x \in \mathbb{R}^n}$$

(normal equation)

Nonlinear case: $f(x) = \sum_{i=1}^{m} r_i(x)^2$.

Try Newton: calculate ∇f, H:

- $$\frac{\partial f}{\partial x_k} = \sum_{i=1}^{m} 2 r_i \frac{\partial r_i}{\partial x_k} = 2\left[\frac{\partial r_1}{\partial x_k} \dots \frac{\partial r_m}{\partial x_k}\right] r$$

Denote $J = \{J_{ij}\} = \left\{\frac{\partial r_i}{\partial x_j}\right\}$ - Jacobian

Then $\boxed{\nabla f = 2 J^T r}$

- $$\frac{\partial^2 f}{\partial x_k \partial x_j} = 2\sum_{i=1}^{m} \left[\frac{\partial r_i}{\partial x_j}\frac{\partial r_i}{\partial x_k} + r_i \frac{\partial^2 r_i}{\partial x_k \partial x_j}\right] \Rightarrow$$

$$\Rightarrow H = 2 J^T J + 2 \sum_{i=1}^{m} r_i \cdot \nabla^2 r_i \approx$$

$$\approx 2 J^T J$$

$$\boxed{x_{k+1} = x_k - (J^TJ)^{-1} J^T r(x_k)}$$

Gauß-Newton

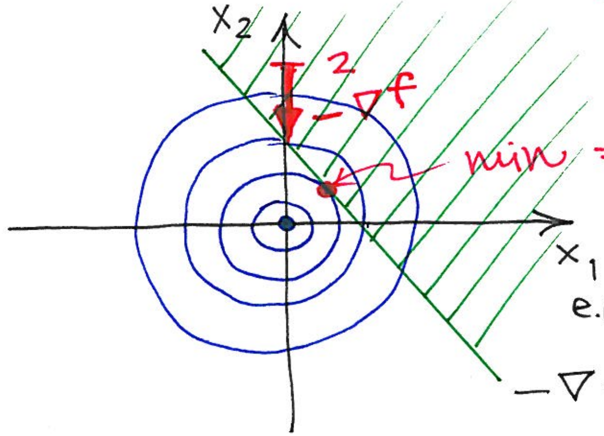
Ch 9. Penalty and Barrier functions

The problem: $\min\limits_{x \in S \subset \mathbb{R}^n} f(x)$

We assumed that $S = \mathbb{R}^n$.

It was crucial in all methods that one can move in any directions.

Ex $\min(x_1^2 + x_2^2) \mid \underbrace{x_1 + x_2 \geq 1}_{\text{constraint}}$



$\text{min} = (\frac{1}{2}, \frac{1}{2})$.

starting at e.g. $(0, 2)$:

$-\nabla f = \begin{bmatrix} 0 \\ -4 \end{bmatrix}$ wrong!

Remark: SD, Newton, CG, Conjugate Directions etc will never find the minimum when applied to $f(x)$.

we need to pass information to the search direction about the constraints.

9.2. Penalty function method

$\min\limits_{x \in S \subset \mathbb{R}^n} f(x)$, $S = \{x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0\}$

Here $g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix}$, $h(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_e(x) \end{bmatrix}$.

$g(x) \leq 0 \iff$ all $g_k(x) \leq 0$.

In theory, it is easy to reduce any S to the case of the whole $\mathbb{R}^n$.

$$F(x) = \begin{cases} f(x) & \text{if } x \in S \\ +\infty & \text{otherwise} \end{cases} \Rightarrow$$

$$\Rightarrow \boxed{\min_{x \in S} f(x) = \min_{x \in \mathbb{R}^n} F(x)}$$

In practice, $+\infty$ is replaced by something large.

For example, take

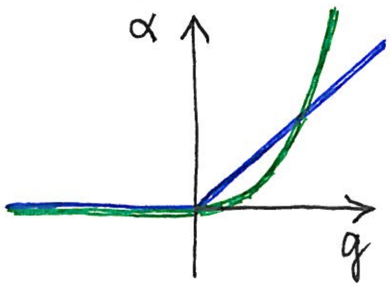
$$\alpha(x) = \begin{cases} 0 & \text{if } x \in S \\ >0 & \text{otherwise} \end{cases} \quad \text{and}$$

build $q(x) = f(x) + \mu \cdot \alpha(x) \Rightarrow$

$\Rightarrow q(x) \approx F(x)$ for large $\mu > 0$.

Typical choice of α:

- $g_k(x) \le 0$:



$\alpha_k(x) = \max\{0, g_k(x)\}$

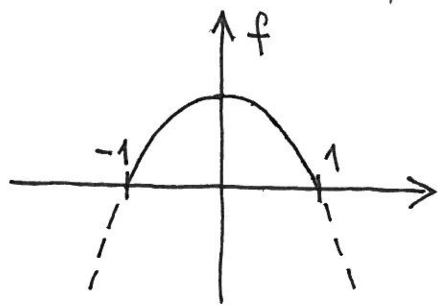
(alt. $\alpha_k(x) = \max\{0, g_k(x)\}^2$)

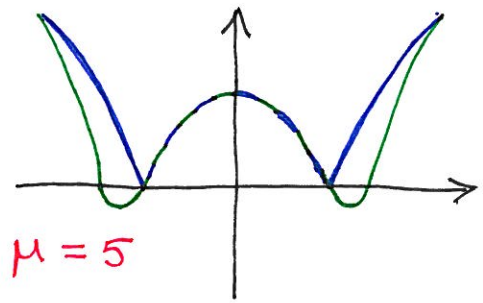
- $h_j(x) = 0$: $\alpha_j(x) = h_j(x)^2$

Overall $\alpha(x)$ for $g(x) \le 0$, $h(x) = 0$:

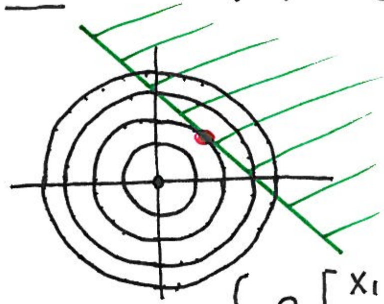
$$\alpha(x) = \sum_{k=1}^{m} \left(\max\{0, g_k(x)\}\right)^2 + \sum_{j=1}^{\ell} h_j(x)^2$$

Ex $f(x) = 1 - x^2$, $-1 \le x \le 1$





Ex $\min(x_1^2 + x_2^2)$, $x_1 + x_2 \ge 1$

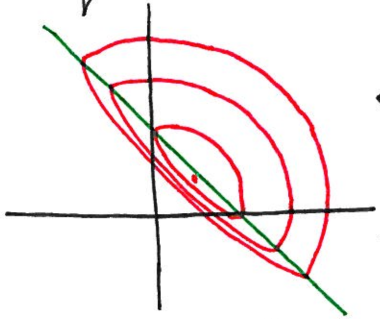


$f(x) = x_1^2 + x_2^2$, $g(x) = 1 - x_1 - x_2$

$$q(x) = \begin{cases} x_1^2 + x_2^2 & \text{if } x_1 + x_2 \ge 1 \\ x_1^2 + x_2^2 + \mu(1 - x_1 - x_2)^2 & \text{otherwise} \end{cases}$$

$$\nabla q(x) = \begin{cases} 2\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} & \text{if } x_1 + x_2 \ge 1 \\ 2\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 2\mu(1 - x_1 - x_2)\begin{bmatrix} -1 \\ -1 \end{bmatrix} & \text{otherwise} \end{cases}$$

$$\nabla q(x) = 0 \iff \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \mu(-1 + x_1 + x_2)\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff$$



$$\iff \begin{bmatrix} 1+\mu & \mu \\ \mu & 1+\mu \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \end{bmatrix} \iff$$

$$\iff \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1+\mu & \mu \\ \mu & 1+\mu \end{bmatrix}^{-1}\begin{bmatrix} \mu \\ \mu \end{bmatrix} =$$

$$= \frac{1}{(1+\mu)^2 - \mu^2}\begin{bmatrix} 1+\mu & -\mu \\ -\mu & 1+\mu \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\mu = \frac{\mu}{1+2\mu}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \xrightarrow[\mu \to \infty]{} \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$$

Remark: • to start with large μ is bad, (ill-conditioned problem). In practice, iterations start from small μ and then gradually increase it using the answers as new starting points.

Strategy: pick $\mu_1 < \mu_2 < \dots < \mu_K \to +\infty$

$$q_\mu(x) = f(x) + \mu \cdot \alpha(x)$$

	μ_1	μ_2	μ_3	...
Start	x_0	x_1	x_2	...
solution to min q_μ	x_1	x_2	x_3	...

• Convergence analysis (Th. 1, p. 316)

$x_K \xrightarrow[\mu\to\infty]{} \bar{x} \Rightarrow \bar{x}$ is the solution to $\min\limits_{x\in S} f(x)$.

• Better to use Newton/quasi Newton/CG

SD is sensitive to ill-conditioned problems.

• Each x_K is not feasible (outside S).

Exterior approximations to $\bar{x}$.

9.3. Barrier function method

Only inequalities: $\min f(x) \mid \underbrace{g(x) \leq 0}_{= S}$

We take another approximation of

$$F(x) = \begin{cases} f(x) & \text{if } x \in S \\ +\infty & \text{otherwise} \end{cases}$$

Take $\beta(x) = \begin{cases} \geq 0 & \text{if } x: g(x) < 0 \\ \to +\infty & \text{if some } g_K(x) \to 0 \\ \text{"}+\infty\text{"}\ ^{*)} & \text{otherwise} \end{cases}$

[*) "$+\infty$" is some very large number to prevent the line search to leave S]

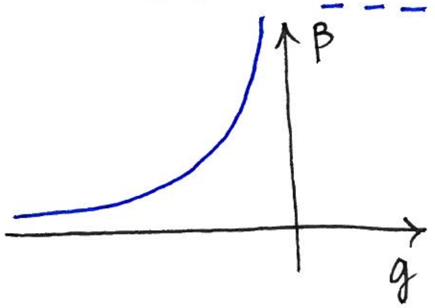
and build $q(x) = f(x) + \varepsilon \cdot \beta(x) \Rightarrow$

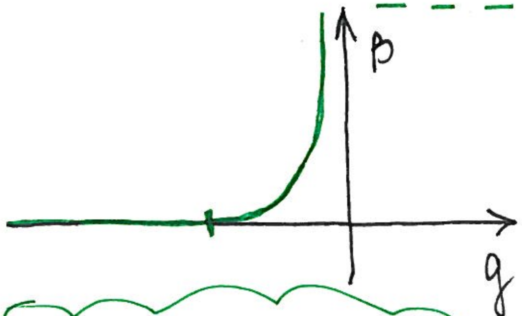
$\Rightarrow q(x) \approx F(x)$ for small $\varepsilon > 0$.

Typical choice of β: for $g_k(x) \leq 0$

- $$\beta_k(x) = \begin{cases} -\dfrac{1}{g_k(x)} & \text{if } g_k(x) < 0 \\ \text{"}+\infty\text{"} & \text{otherwise} \end{cases}$$

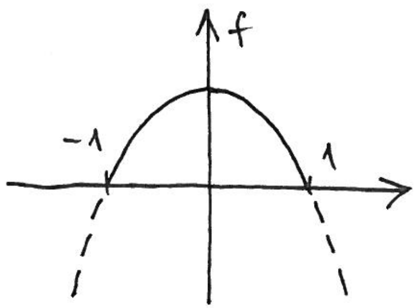
- $$\beta_k(x) = \begin{cases} 0 & \text{if } g_k(x) \leq -1 \\ -\ln(-g_k(x)) - g_k(x) - 1 & \text{if } -1 < g_k(x) < 0 \\ \text{"}+\infty\text{"} & \text{otherwise} \end{cases}$$

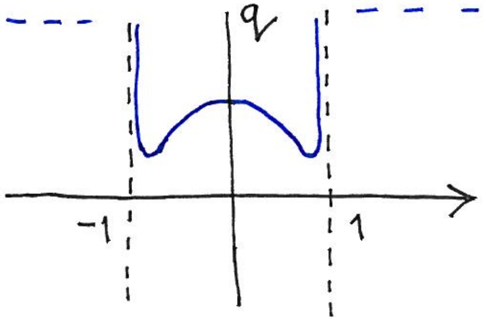




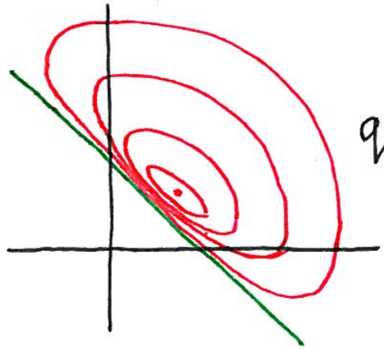
General $g(x) \leq 0$: $\boxed{\beta(x) = \sum_{k=1}^{m} \beta_k(x)}$

Ex $f(x) = 1 - x^2$, $-1 \leq x \leq 1$





Ex (again) $\min(x_1^2 + x_2^2) \mid x_1 + x_2 \geq 1$



$$q(x) = \begin{cases} x_1^2 + x_2^2 + \dfrac{\varepsilon}{x_1 + x_2 - 1} & \text{if } x_1 + x_2 > 1 \\ 10^{50} & \text{otherwise} \end{cases}$$

(alt. realmax in MATLAB)

Remark:
- similar strategy as for penalty:

Pick not very small $\varepsilon_1 > \varepsilon_2 > \ldots > \varepsilon_k \to 0^+$

$$q_\varepsilon(x) = f(x) + \varepsilon\, \beta(x)$$

	ε_1	ε_2	ε_3	...
start	x_0	x_1	x_2	...
solution to $\min q_\varepsilon$	x_1	x_2	x_3	...

- Similar convergence analysis (Th. 2, p. 325)
- SD is no good here either.
- Line search must be used to stay in S.
- Each x_k is feasible.

Interior approximations to $\bar{x}$.