Optimization: Nonlinear Optimization with Constraints

Constraints in nonlinear optimization

Equality constraints

- Linear equality constraints
- Nonlinear equality constraints

Inequality constraints

- Linear inequality constraints
- Nonlinear inequality constraints
  - penalty/barrier function
  - SQP: Sequential Quadratic Programming

Equality constraints

Linear constraints → Elimination

$$\min_{x \in \mathbb{R}^n} f(x)\,, \quad \text{where} \;\; Ax = b$$

$$x = x_0 + \bar{A}^T \bar{x}$$

such that $Ax_0 = b$ and $A\bar{A}^T = 0$

$$\min_{\bar{x} \in \mathbb{R}^{(n-m)}} f(x_0 + \bar{A}^T \bar{x})$$

Use SVD (Singular Value Decomposition):

$$A = U \begin{bmatrix} \Sigma & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U \, \Sigma \, V_1^T$$

Define $\bar{A} = V_2^T$ and $x_0 = V_1 \, \Sigma^{-1} \, U^T \, b$

Equality constraints

Nonlinear equality constraints

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \;\; h(x) = 0$$

Consider the function $f(x) + \lambda^T h(x)$

Note that zero-gradient condition for $f(x) + \lambda^T h(x)$, i.e.,

$$\nabla_{x,\lambda} \Big(f(x) + \lambda^T h(x) \Big) = 0$$

is equivalent to Lagrange conditions

$$\nabla_x f(x) + \nabla_x h(x)\, \lambda = 0$$

$$\nabla_\lambda \Big(\lambda^T h(x) \Big) = h(x) = 0$$

Equality constraints

Unfortunately, it can be shown that local minima of $\min_x f(x)$ s.t. $h(x) = 0$ correspond to *saddle* points of $f(x) + \lambda^T h(x)$

So solving $\min_{x,\lambda} f(x) + \lambda^T h(x)$ does not work

However, to obtain points that satisfy zero-gradient condition

$$\nabla_{x,\lambda} \left(f(x) + \lambda^T h(x) \right) = 0$$

we can equivalently solve

$$\min_{x,\lambda} \left\| \nabla_{x,\lambda} (f(x) + \lambda^T h(x)) \right\|_2^2$$

→ unconstrained optimization problem!

Inequality constraints

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leqslant 0$$

Elimination

Mapping $\Phi : \bar{x} \to x$ such that

$$\{\, x = \Phi(\bar{x}),\ \bar{x} \in \mathbb{R}^m \,\} = \{\, x \mid x \in \mathbb{R}^n \,,\ g(x) \leqslant 0 \,\}$$

New unconstrained minimization problem

$$\min_{\bar{x} \in \mathbb{R}^m} f\Big(\Phi(\bar{x})\Big)$$

Gradient projection method

Linear inequality constraints: $\min_{x \in \mathbb{R}^n} \; f(x)$ s.t. $A\,x - b \leqslant 0$

What if $-\nabla f(x_k)$ points outside feasible region in boundary point x_k?

For boundary point x_k: $a_j^T x_k = b_j$ for $j \in \mathcal{A}$ $\rightarrow$ "active"

$a_j^T x_k < b_j$ for $j \notin \mathcal{A}$

Rows indexed by $\mathcal{A} \rightarrow$ submatrices A_a and b_a with

$$A_a\, x_i = b_a$$

Define projection matrix:

$$P = I - A_a^T (A_a\, A_a^T)^{-1} A_a$$

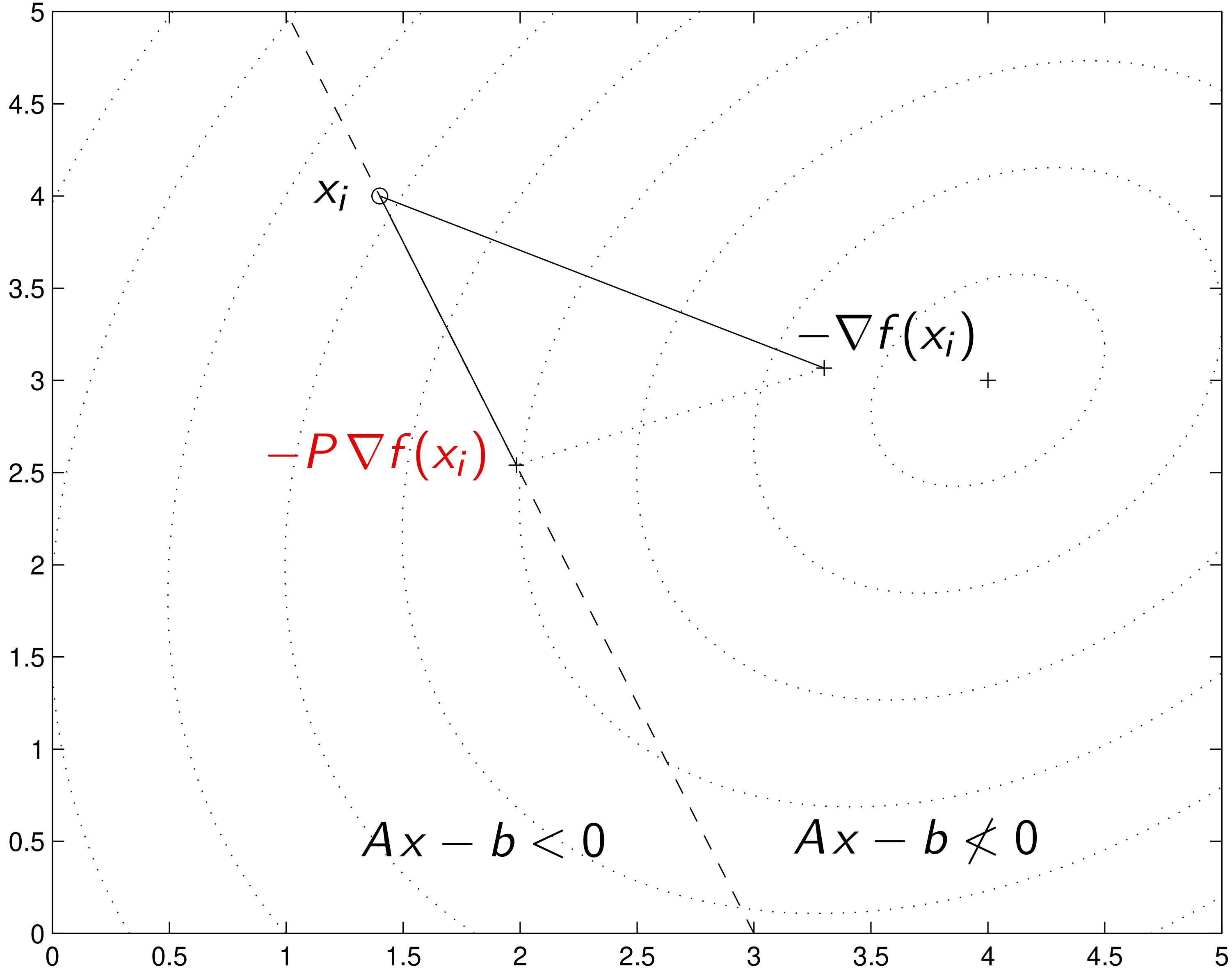
New search direction:

$$d_k = -P\, \nabla f(x_k)$$

One-dimensional minimization problem:

$$\min_{s \in \mathbb{R}} f(x_k + d_k\, s) \qquad \text{s.t.} \quad A\,(x_k + d_k\, s) - b \leqslant 0$$

Gradient projection method (continued)



Inequality constraints — Penalty/barrier function

Nonlinear inequality constraints

$$\min_{x \in \mathbb{R}^n} f(x)\,, \quad \text{where} \;\; g(x) \leqslant 0$$

Ideally: *feasibility* function $f_{\text{feas}}(x)$ given by

$$\begin{aligned} f_{\text{feas}}(x) &= 0 && \text{if} \;\; \max_i g_i(x) \leqslant 0 \quad (\text{or: } \; g(x) \leqslant 0) \\ f_{\text{feas}}(x) &= \infty && \text{if} \;\; \max_i g_i(x) > 0 \quad (\text{or: } \; g(x) \not\leqslant 0) \end{aligned}$$

Unconstrained minimization:

$$\min_x \Big(\, f(x) + f_{\text{feas}}(x) \, \Big)$$

Feasibility function is not smooth !!

- Penalty function
- Barrier function

Penalty function

$$f_{\text{pen}}(x) = 0 \quad \text{for} \quad \max_i g_i(x) \leqslant 0$$

$$f_{\text{pen}}(x) \gg 0 \quad \text{for} \quad \max_i g_i(x) > 0$$

Examples of penalty functions are:

$$f_{\text{pen}} = \beta \sum_{i=1}^{m} \max\Big(0, g_i(x)\Big) \quad , \quad \beta \gg 1$$

$$f_{\text{pen}} = \beta \sum_{i=1}^{m} \max\Big(0, g_i(x)\Big)^2 \quad , \quad \beta \gg 1$$

$$f_{\text{pen}} = \max_i \max(0, e^{\beta\, g_i(x)} - 1)^2 \quad , \quad \beta \gg 1$$

Barrier function

$$f_{\text{bar}}(x) \approx 0 \quad \text{for} \quad \max_i g_i(x) \ll 0$$

$$f_{\text{bar}}(x) \longrightarrow \infty \quad \text{for} \quad \max_i g_i(x) \uparrow 0$$

usually undefined for $\max_i g_i(x) \geqslant 0$

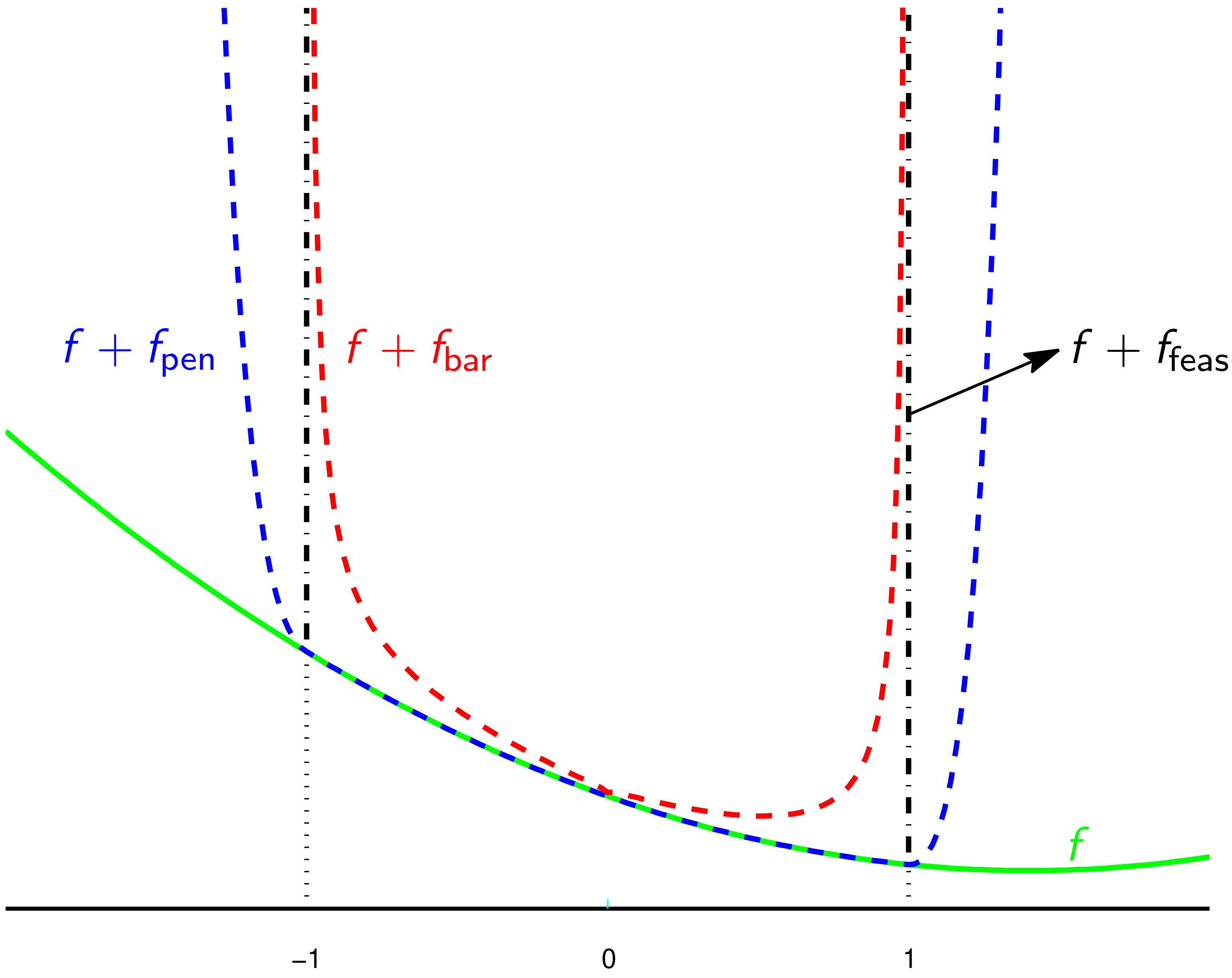
Examples of barrier functions are:

$$f_{\text{bar}} = -\frac{1}{\beta} \sum_{i=1}^{m} \ln\Big(-g_i(x)\Big) \ , \quad \beta > 1$$

$$f_{\text{bar}} = -\frac{1}{\beta} \sum_{i=1}^{m} \frac{1}{g_i(x)} \ , \quad \beta > 1$$

$$f_{\text{bar}} = -\frac{1}{\beta} \ln\Big(-\max_i g_i(x)\Big) \ , \quad \beta > 1$$

Penalty & barrier functions



Sequential Quadratic Programming

State-of-the art algorithm for

$$\min_x \; f(x) \quad \text{s.t.} \quad g(x) \leqslant 0$$

Idea 1:
approximate f by quadratic function, g by linear function
$\rightarrow$ does not always work in practice

Idea 2:
use Lagrange function:

$$L(x,\lambda) = f(x) + \lambda^T g(x)$$

$$\Rightarrow \quad \min_x \; L(x,\lambda) \quad \text{s.t.} \quad g(x) \leqslant 0$$

zero-gradient condition: $\nabla_x L(x,\lambda) = 0$

$=$

first Karush-Kuhn-Tucker condition: $\nabla f(x) + \lambda^T \nabla g(x) = 0$

SQP (continued)

Quadratic approximation for L:

$$L(x, \lambda_k) \approx L(x_k, \lambda_k) + \nabla_x^T L(x_k, \lambda_k) \underbrace{(x - x_k)}_{d} +$$

$$\frac{1}{2} \underbrace{(x - x_k)^T}_{d^T} H_L(x_k, \lambda_k) \underbrace{(x - x_k)}_{d}$$

Linear approximation of g:

$$g(x) = g(x_k) + \nabla^T g(x_k) \underbrace{(x - x_k)}_{d}$$

$\rightarrow$ quadratic programming problem in d

Note: In literature $\nabla f(x_k)$ is mostly used instead of $\nabla_x L(x_k, \lambda_k)$ in quadratic objective function since this yields better performance

SQP algorithm

1. Current point: x_k, λ_k
2. Compute (approximations) of $\nabla f(x_k)$ and $H_L(x_k, \lambda_k)$: G_k, H_k
3. Define $d = x - x_k$ and solve QP:

$$\min_d \; \frac{1}{2} d^T H_k d + G_k^T d$$

$$\text{s.t. } \; g(x_k) + \nabla^T g(x_k) \, d \leqslant 0$$

   $\Rightarrow d_k = d^*$, $\Delta_k = \lambda^* - \lambda_k$
   with λ^* the optimal Lagrange multiplier for the QP
4. Perform line search: $s_k = \arg\min_s \; \psi(x_k + s\, d_k)$
   with, e.g., $\psi = f + f_{\text{pen}}$
5. Define the new estimate: $x_{k+1} = x_k + s_k\, d_k$

$$\lambda_{k+1} = \lambda_k + s_k\, \Delta_k$$

6. If not optimal, goto step 1.

Summary

- Nonlinear optimization with constraints: Standard form

$$\begin{aligned} \min_x \; & f(x) \\ \text{s.t.} \; h(x) &= 0 \\ g(x) &\leqslant 0 \end{aligned}$$

- Main solution approaches:
  - Elimination of constraints!
  - Nonlinear equality constraints $\rightarrow$ Lagrange
  - Linear inequality constraints $\rightarrow$ gradient projection
  - Nonlinear inequality constraints $\rightarrow$ penalty or barrier function, SQP