Numerical Optimization

Unit 10: Quadratic Programming, Active Set Method, and Sequential Quadratic Programming

Che-Rung Lee

Department of Computer Science
National Tsing Hua University

November 17, 2021

Quadratic programming

The General form

$$\min_{\vec{x}} g(\vec{x}) = \frac{1}{2}\vec{x}^T G\vec{x} + \vec{x}^T\vec{c}$$

$$\begin{aligned} s.t. \quad & \vec{a}_i^T\vec{x} = b_i \quad i \in \mathcal{E} \\ & \vec{a}_i^T\vec{x} \geq b_i \quad i \in \mathcal{I} \end{aligned}$$

The Lagrangian

$$\mathcal{L}(\vec{x},\vec{\lambda}) = \frac{1}{2}\vec{x}^T G\vec{x} + \vec{x}^T\vec{c} - \vec{\lambda}^T(A\vec{x} - \vec{b})$$

$$A = \begin{bmatrix} \vec{a}_1^T \\ \vec{a}_2^T \\ \vdots \\ \vec{a}_m^T \end{bmatrix} \in \mathbb{R}^{m\times n} \quad (\text{assuming } m \leq n)$$

KKT Condition

KKT condition

$$\nabla \mathcal{L}(\vec{x}, \vec{\lambda}) = 0$$

$$\vec{a}_i^T \vec{x} = b_i \quad i \in \mathcal{E}$$

$$\vec{a}_i^T \vec{x} \geq b_i \quad i \in \mathcal{I}$$

$$\lambda_i \geq 0 \quad i \in \mathcal{I}$$

$$\lambda_i(\vec{a}_i^T \vec{x} - \vec{b}_i) = 0, \quad i \in \mathcal{I}$$

1. If G is positive definite and $\vec{x}^*, \vec{\lambda}^*$ satisfy KKT conditions, then $\vec{x}^*$ is the global solution of the optimization problem.

KKT Matrix

- Let's first consider the equality constraints only

$$\nabla \mathcal{L}(\vec{x}, \vec{\lambda}) = 0 \Rightarrow \begin{aligned} G\vec{x} - A^T\vec{\lambda} &= -\vec{c} \\ A\vec{x} &= \vec{b} \end{aligned}$$

$$\Rightarrow \begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \vec{x} \\ \vec{\lambda} \end{bmatrix} = \begin{bmatrix} -\vec{c} \\ \vec{b} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \vec{x} \\ -\vec{\lambda} \end{bmatrix} = \begin{bmatrix} -\vec{c} \\ \vec{b} \end{bmatrix} \tag{1}$$

- The matrix $\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix}$ is called the KKT matrix.
- If A has full row-rank and the reduced Hessian Z^TGZ is positive definite, where span$\{Z\}$ is the null space of span$\{A^T\}$ then the KKT matrix is nonsingular.
- If there are only equality constraints, solve (1) directly can get optimal solution.

Example

$$\min_{\vec{x}} \vec{x}^T \begin{bmatrix} 6 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{bmatrix} \vec{x} + \begin{bmatrix} -8 \\ -8 \\ -3 \end{bmatrix}^T \vec{x} \text{ s.t. } \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \vec{x} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

- The optimal solution is at $\vec{x}^* = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix}^T$, $\vec{\lambda}^* = [3, -2]^T$.
- The optimal solution can be obtained by solving

$$\left[\begin{array}{cc} G & A^T \\ A & 0 \end{array} \right] \left[\begin{array}{c} \vec{x} \\ -\vec{\lambda} \end{array} \right] = \left[\begin{array}{c} -\vec{c} \\ \vec{b} \end{array} \right]$$

- The KKT matrix and the right hand side

$$\begin{bmatrix} G & A^T \\ A & O \end{bmatrix} = \begin{bmatrix} 6 & 2 & 1 & 1 & 0 \\ 2 & 5 & 2 & 0 & 1 \\ 1 & 2 & 4 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}, \vec{c} = \begin{bmatrix} -8 \\ -8 \\ -3 \end{bmatrix}, \vec{b} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

Variable elimination 1/2

A general strategy for linear equality constraints is variable elimination.

Variable elimination

- Let $A\vec{x} = \vec{b}$ be the linear equality constraints. $A \in \mathbb{R}^{m\times n}, \vec{x} \in \mathbb{R}^n$, and $\vec{b} \in \mathbb{R}^m$. we assume $m < n$.
- We can choose m linearly independent columns to be "basic variables" and use them to solve the constraints. Others are called "nonbasic variables", setting to 0. Let

$$AP = [B|N], \left(\begin{array}{c} \vec{x}_B \\ \vec{x}_N \end{array} \right) = P^T\vec{x}$$

where P is a permutation matrix.

$$\vec{b} = A\vec{x} = APP^T\vec{x} = B\vec{x}_B + N\vec{x}_N.$$

Variable elimination 2/2

- Therefore, $\vec{x}_B = B^{-1}\vec{b} - B^{-1}N\vec{x}_N$. The original constrained problem becomes an unconstrained problem

$$\begin{array}{ll} \min_{\vec{x}} & f(\vec{x}) \\ \text{s.t.} & A\vec{x} = \vec{b} \end{array} \implies \min_{\vec{x}_N} f\left(\left[\begin{array}{c} B^{-1}\vec{b} - B^{-1}N\vec{x}_N \\ \vec{x}_N \end{array}\right]\right)$$

- For nonlinear equality constraints, variable elimination may not feasible.
- For example,

$$\min_{x,y} x^2 + y^2$$

$$\text{s.t. } (x-1)^3 = y^2$$

The solution is at $(x, y) = (1, 0)$. Using variable elimination, the problems becomes $\min_x x^2 + (x-1)^3$ which is unbounded.

Schur complement method

- The block LU decomposition of $\begin{bmatrix} G & A^T \\ A & O \end{bmatrix}$ is

$$\begin{bmatrix} G & A^T \\ A & O \end{bmatrix} = \begin{bmatrix} I & O \\ AG^{-1} & I \end{bmatrix} \begin{bmatrix} G & A^T \\ O & -AG^{-1}A^T \end{bmatrix}$$

  where $-AG^{-1}A^T$ is called the Schur complement.
- The inverse of $\begin{bmatrix} I & O \\ AG^{-1} & I \end{bmatrix}$ is $\begin{bmatrix} I & 0 \\ -AG^{-1} & I \end{bmatrix}$.
  So the KKT system can be rewritten as

$$\begin{bmatrix} G & A^T \\ O & -AG^{-1}A^T \end{bmatrix} \begin{bmatrix} \vec{x} \\ -\vec{\lambda} \end{bmatrix} = \begin{bmatrix} I & 0 \\ -AG^{-1} & I \end{bmatrix} \begin{bmatrix} -\vec{c} \\ \vec{b} \end{bmatrix} = \begin{bmatrix} -\vec{c} \\ AG^{-1}\vec{c} + \vec{b} \end{bmatrix}$$

Example

$$\min_{\vec{x}} \vec{x}^T \begin{bmatrix} 6 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{bmatrix} \vec{x} + \begin{bmatrix} -8 \\ -8 \\ -3 \end{bmatrix}^T \vec{x} \text{ s.t. } \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \vec{x} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

- The optimal solution is at $\vec{x}^* = \begin{bmatrix} 2 & -1 & 1 \end{bmatrix}^T$, $\vec{\lambda}^* = \begin{bmatrix} 3, -2 \end{bmatrix}^T$.
- The optimal solution can be obtained by solving

$$\left[\begin{array}{cc} G & A^T \\ A & 0 \end{array}\right] \left[\begin{array}{c} \vec{x} \\ -\vec{\lambda} \end{array}\right] = \left[\begin{array}{c} -\vec{c} \\ \vec{b} \end{array}\right]$$

- The KKT matrix and the right hand side

$$\begin{bmatrix} G & A^T \\ A & O \end{bmatrix} = \begin{bmatrix} 6 & 2 & 1 & 1 & 0 \\ 2 & 5 & 2 & 0 & 1 \\ 1 & 2 & 4 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}, \vec{c} = \begin{bmatrix} -8 \\ -8 \\ -3 \end{bmatrix}, \vec{b} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

Example: Schur complement method

The equations from Schur complement are

$$G\vec{x} - A^T\vec{\lambda} = -\vec{c}$$
$$AG^{-1}A^T\vec{\lambda} = AG^{-1}\vec{c} + \vec{b}$$

$$AG^{-1}A^T = \begin{bmatrix} 0.4819 & 0.1084 \\ 0.1084 & 0.3494 \end{bmatrix}$$

$$\vec{\lambda} = (AG^{-1}A^T)^{-1}(AG^{-1}\vec{c} + \vec{b}) == \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

$$\vec{x} = G^{-1}(A^T\vec{\lambda} + \vec{c}) = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

Active set method

Active set method

Active set method solves constrained optimization problems by searching solutions in the feasible sets.

- If constraints are linear and one can guess the active constrains for the optimal solution, then one can use the active constraints to reduce the number of unknowns, and then perform algorithms for unconstrained optimization problems.
- Problem: how to guess the set of active constraints.
- Linear programming is an active set method.

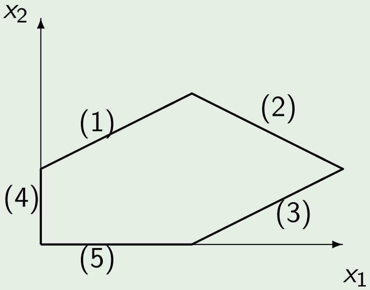
Active set method for convex QP

Consider the following example

Example

$\min_x g(\vec{x}) = (x_1 - 1)^2 + (x_2 - 2.5)^2$

$$
\begin{aligned}
\text{s.t.} \qquad x_1 - 2x_2 + 2 &\geq 0 \quad &&\text{—(1)}\\
-x_1 - 2x_2 + 6 &\geq 0 \quad &&\text{—(2)}\\
-x_1 + 2x_2 + 2 &\geq 0 \quad &&\text{—(3)}\\
x_1, x_2 &\geq 0 \quad &&\text{—(4),(5)}
\end{aligned}
$$



- Initial step $\vec{x}_0 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$
- The working set $\mathcal{W}_0 = \{(3), (5)\}$

Solve the EQP

Example

$$\min_{\vec{x}} g(\vec{x}) = x_1^2 - 2x_1 + 1 + x_2^2 - 5x_2 + \frac{25}{4}$$

$$\min_{\vec{x}} \vec{x} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \vec{x} + \begin{pmatrix} -2 \\ -5 \end{pmatrix}^T \vec{x} + \frac{29}{4}$$

$$\text{s.t.} \quad \begin{pmatrix} -1 & 2 \\ 0 & 1 \end{pmatrix} \vec{x} = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

Since it has equality constraints only, using KKT system to solve the QP.

$$K = \begin{pmatrix} 2 & 0 & -1 & 0 \\ 0 & 2 & 2 & 1 \\ -1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \vec{x} \\ -\vec{\lambda} \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ -2 \\ 0 \end{pmatrix}$$

Optimality check

Example

- The solution of the KKT system is $\vec{x}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \vec{\lambda}_1 = \begin{pmatrix} -2 \\ -1 \end{pmatrix}$
- Both Lagrangian multipliers are negative
  $\Rightarrow$ This is not the optimal solution.
  $\Rightarrow$ Remove one of the constraint $\mathcal{W}_1 = \{(5)\}$ and solve the KKT system again.

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \vec{x}_2 \\ -\vec{\lambda}_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 0 \end{pmatrix}, \vec{x}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \vec{\lambda}_2 = -5$$

- Let $\vec{p}_1 = \vec{x}_2 - \vec{x}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ be the search direction, and search $\alpha_1 \in [0, 1]$, such that $\vec{x}_2^+ = \vec{x}_1 + \alpha_1 \vec{p}_1$ is feasible.

Feasibility check

Example

- The feasibility check: $\vec{x}_1 + \alpha_1 \vec{p}_1 = \begin{pmatrix} 2-\alpha_1 \\ 0 \end{pmatrix} \Rightarrow \alpha_1 = 1$

  Move to $\vec{x}_2 = \vec{x}_2^+ = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

- But $\vec{\lambda}_2 < 0$, it is not the optimal solution.
  $\Rightarrow$ Remove one more constraint $\mathcal{W}_2 = \emptyset$

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \left(\vec{x}_3 \right) = \begin{pmatrix} 2 \\ 5 \end{pmatrix} \Rightarrow \vec{x}_3 = \begin{pmatrix} 1 \\ 2.5 \end{pmatrix}$$

Let $\vec{p}_2 = \begin{pmatrix} 1 \\ 2.5 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 2.5 \end{pmatrix}$ be the search direction.

$$\vec{x}_3^+ = \vec{x}_2 + \alpha_2 \vec{p}_2 = \begin{pmatrix} 1 \\ 2.5\alpha_2 \end{pmatrix}, \quad \alpha_2 \in [0,\ 1]$$

Adding new constraints

Example

- For $\alpha_2 = 1$, constraint (1) will be invalided:
  $\Rightarrow x_1 - 2x_2 + 2 \geq 0 \Rightarrow \alpha_2 \leq 0.6$.
- Move to $\vec{x}_3 = \begin{pmatrix} 1 \\ \frac{5}{2} \cdot 0.6 \end{pmatrix} = \begin{pmatrix} 1 \\ 1.5 \end{pmatrix}$, and add (1) to the working set, $\mathcal{W}_3 = \{(1)\}$.
- Solve the KKT conditions:

$$\begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & -2 \\ 1 & -2 & 0 \end{pmatrix} \begin{pmatrix} \vec{x}_4 \\ -\vec{\lambda}_4 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ -2 \end{pmatrix} \qquad \begin{array}{l} \vec{x}_4 = \begin{pmatrix} 1.4 \\ 1.7 \end{pmatrix} \\ \vec{\lambda}_4 = 0.8 \end{array}$$

Since $\lambda \geq 0$ and all constraints are satisfied, it is the optimal solution.

Gradient projection method

A special case of inequality constrains are bounded constraints

$$\min_{\vec{x}} q(\vec{x}) = \frac{1}{2}\vec{x}^T G\vec{x} + \vec{x}^T\vec{c}$$

$$\text{s.t.} \quad \vec{l} \leq \vec{x} \leq \vec{u} \quad (\text{which means } l_i \leq x_i \leq u_i \text{ forall } i)$$

which can be solved by gradient project method.

Algorithm: Gradient projection method

1. Given $\vec{x}_0$.
2. For $k = 0, 1, 2, \ldots$ until converge
   (a) Find a search direction $\vec{g}$.
   (b) Construct a piece wise linear function $x(t) = p(\vec{x} - t\vec{g}, \vec{l}, \vec{u})$
   (c) In each line sequent of $x(t)$ find the optimal solution $\vec{x}^c$
   (d) Use $\vec{x}^c$ as an initial guess to solve $\min_{\vec{x}} q(\vec{x})$

$$\text{s.t.} \begin{cases} x_i = x_i^c & i \in A(\vec{x}^c) \\ l_i \leq x_i \leq u_i & i \notin A(\vec{x}^c) \end{cases}$$

In the algorithm 1/2

- For 2(a), the search direction can be any descent direction, such as $-\nabla q$.
- For 2(b), the piecewise linear function is computed as

$$p(\vec{x}, \vec{l}, \vec{u})_i = \begin{cases} l_i & \text{if } x_i < l_i \\ x_i & \text{if } x_i \in [l_i, u_i] \\ u_i & \text{if } x_i > u_i \end{cases}$$

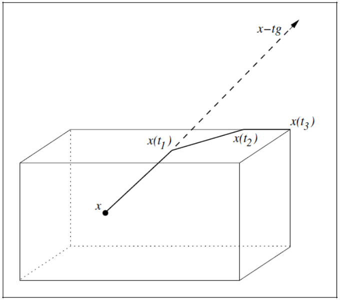


Figure 16.4 The piecewise-linear path $x(t)$, for an example in $\mathbb{R}^3$.

- For each element x_i, compute $\bar{t}_i$ as

$$\bar{t}_i = \begin{cases} (l_i - x_i)/g_i & \text{if } g_i < 0, \quad \text{and } l_i > -\infty \\ (u_i - x_i)/g_i & \text{if } g_i > 0, \quad \text{and } u_i < +\infty \\ \infty & \text{otherwise} \end{cases}$$

- Sort $\{\bar{t}_1, \bar{t}_2, ..., \bar{t}_n\}$ to get $t_0 = 0 < t_1 < t_2 < ... < t_{m-1} < \infty = t_m$

In the algorithm 2/2

- For each $[t_{j-1}, t_j]$, $x(t) = x(t_{j-1}) + (t - t_{j-1})\vec{p}^{j-1}$, where

$$p_i^{j-1} = \begin{cases} -g_i & \text{if } t_{j-1} < \bar{t}_i \\ 0 & \text{otherwise} \end{cases}$$

- For 2(c), we search optimal solution segment by segment $[t_{j-1}, t_j]$.
- Let $\Delta t = t - t_{j-1}, x(\Delta t) = x(t_{j-1}) + \Delta t \vec{p}^{j-1}$.

$$\begin{aligned} q(x(\Delta t)) &= \vec{c}^T(x(t_{j-1}) + \Delta t \vec{p}^{j-1}) + \\ & \quad \frac{1}{2}(x(t_{j-1} + \Delta t \vec{p}^{j-1})^T G(x(t_{j-1} + \Delta t \vec{p}^{j-1})) \\ &= \frac{1}{2}a(\vec{p}^{j-1})\Delta t^2 + b(\vec{p}^{j-1})\Delta t + c(\vec{p}^{j-1}) \end{aligned}$$

for some function a, b, c of $\vec{p}^{j-1}$.

Optimal $\Delta t^* = \dfrac{-b(\vec{p}^{j-1})}{a(\vec{p}^{j-1})}$ and $t^* = t_{j-1} + \Delta t^*$.

Example

$$\min_{x_1,x_2,x_3} f = (x_1+2x_2-1)^2+(x_2+2x_3-2)^2+(x_3-3/4)^2 \text{ s.t. } 0 \leq x_1, x_2, x_3 \leq 1.$$

At which point the optimal solution is?

- In the quadratic model,

$$\min_{\vec{x}} f(\vec{x}) = \frac{1}{2}\vec{x}^T G\vec{x} + \vec{c}^T\vec{x} + d$$

$$G = \begin{bmatrix} 2 & 4 & 0 \\ 4 & 10 & 4 \\ 0 & 4 & 10 \end{bmatrix}, \vec{c} = \begin{bmatrix} -2 \\ -6 \\ -11/2 \end{bmatrix}, d = \frac{89}{16}.$$

- Gradient

$$\vec{g} = G\vec{x} + \vec{c} = \begin{bmatrix} 2x_1 + 4x_2 - 2 \\ 4x_1 + 10x_2 + 4x_3 - 8 \\ 4x_2 + 10x_3 - 19/2 \end{bmatrix}$$

Example—continue

The initial guess is

$$\vec{x}_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \vec{g}_0 = \begin{bmatrix} -2 \\ -8 \\ -19/2 \end{bmatrix}, \text{ solve } t_i = \begin{cases} (x_i - u_i)/g_i, & \text{if } g_i < 0 \\ (x_i - l_i)/g_i, & \text{if } g_i > 0 \\ \infty, & \text{otherwise} \end{cases}$$

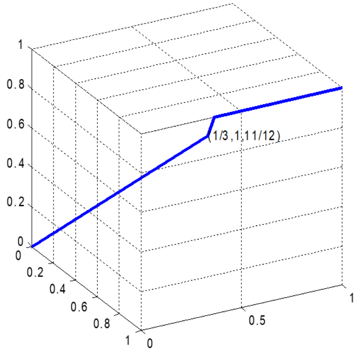
The solutions are $t_1 = 1/2, t_2 = 1/8, t_3 = 2/19$. Since $t_3 < t_2 < t_1$, GP will process the segment, $(0, t_3), (t_3, t_2), (t_2, t_1)$, one by one, using

$$x_i(t) = \begin{cases} x_i - tg_i, & \text{if } t < t_i \\ x_i - t_i g_i, & \text{otherwise} \end{cases}$$

$$x(t) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \to \begin{bmatrix} 2t \\ 8t \\ 19/2t \end{bmatrix} \to \begin{bmatrix} 2t \\ 8t \\ 1 \end{bmatrix} \to \begin{bmatrix} 2t \\ 1 \\ 1 \end{bmatrix} \to \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Example—continue

- For $0 < t < 2/19$, the minimizer is at $\Delta t = 0.0678$. $x_1^* = [0.1356\ 0.5424\ 0.6441]^T$, and $f(x_1^*) = 0.08848$.
- For $2/19 < t < 1/8$, the minimizer is at $\Delta t = -8/13 < 2/19$. So $x_2^* = [4/19, 16/19, 1]^T$ and $f(x_2^*) = 1.572$.
- For $1/8 < t < 1/2$, the minimizer is at $\Delta t = -1/2 < 1/8$. So $x_3^* = [1/4, 1, 1]^T$ and $f(x_3^*) = 2.625$.



- Since the smallest value among all segments is at $t = 0.0682$, the solution obtained by gradient project is x_1^*, and $f(x_1^*) = 0.08848$.
- The exact solution is $x_1^* = [0\ 1/2\ 3/4]^T$, and $f^* = 0$.

Sequential quadratic programming

- Recall the Newton's method for unconstrained problem. It builds a quadratic model at each x_K and solve the quadratic problem at every step.
- SQP uses similar idea: It builds a QP at each step,
  $f : \mathbb{R}^n \to \mathbb{R}, \qquad c : \mathbb{R}^n \to \mathbb{R}^m$

$$\min_{\vec{x}} f(\vec{x}) \qquad s.t. \quad c(\vec{x}) = 0$$

- Let $A(\vec{x})$ be the Jacobian of $c(\vec{x})$:
  $A(\vec{x}) = \begin{pmatrix} \nabla c_1 & \nabla c_2 & \cdots & \nabla c_m \end{pmatrix}^T$
- The Lagrangian $\mathcal{L}(\vec{x}, \vec{\lambda}) = f(\vec{x}^*) - \vec{\lambda}^T c(\vec{x})$
- The KKT condition: $\nabla f(\vec{x}^*) - A(\vec{x}^*)^T \vec{\lambda}^* = 0$, $c(\vec{x}^*) = 0$

Newton's method

- Let $F(\vec{x}, \vec{\lambda}) = \begin{bmatrix} \nabla_x \mathcal{L}(\vec{x}, \vec{\lambda}) \\ c(\vec{x}) \end{bmatrix} = \begin{bmatrix} \nabla f(\vec{x}) - A(\vec{x})^T \vec{\lambda} \\ c(\vec{x}) \end{bmatrix}$. The optimal solution $\vec{x}^*, \vec{\lambda}^*$ must satisfy the KKT condition $\Rightarrow F(\vec{x}^*, \vec{\lambda}^*) = 0$.
- Using Newton's method to solve $F = 0$.
- The Jacobian $\nabla F(\vec{x}, \vec{\lambda}) = \begin{bmatrix} \nabla^2_{xx} \mathcal{L} & -A(\vec{x})^T \\ A(\vec{x}) & 0 \end{bmatrix}$
- The Newton step

$$\begin{bmatrix} \vec{x}_{k+1} \\ \vec{\lambda}_{k+1} \end{bmatrix} = \begin{bmatrix} \vec{x}_k \\ \vec{\lambda}_k \end{bmatrix} + \begin{bmatrix} \vec{p}_k \\ \vec{\ell}_k \end{bmatrix},$$

where

$$\begin{bmatrix} \nabla^2_{xx} \mathcal{L} & -A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} \vec{p}_k \\ \vec{\ell}_k \end{bmatrix} = \begin{bmatrix} -\nabla f_k + A_k^T \vec{\lambda}_k \\ -\vec{c}_k \end{bmatrix} \tag{2}$$

(We use $A_k = A(\vec{x}_k)$, $f_k = f(\vec{x}_k)$, and $\vec{c}_k = c(\vec{x}_k)$.)

Alternative formulation

- To simply that, we examine the first equation

$$\nabla^2_{xx}\mathcal{L}\vec{p}_k - A_k^T\vec{\ell}_k = -\nabla f_k + A_k^T\vec{\lambda}_k$$

- Since $A(x_k)^T(\vec{\lambda}_k + \vec{\ell}_k) = A(x_k)^T\vec{\lambda}_{k+1}$, we can rewrite (2) as

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L} & A_k^T \\ A_k & 0 \end{bmatrix}\begin{bmatrix} \vec{p}_k \\ -\vec{\lambda}_{k+1} \end{bmatrix} = \begin{bmatrix} -\nabla f_k \\ -\vec{c}_k \end{bmatrix} \tag{3}$$

- If A_k is of full row-rank and $Z^T\nabla^2_{xx}\mathcal{L}Z$ is positive definition, where Z is the null space of spanA_k, then the above equation solves the following QP

$$\min_{\vec{p}} \frac{1}{2}\vec{p}^T\nabla^2_{xx}\mathcal{L}\vec{p} + \nabla_x f_k^T\vec{p}$$

$$\text{s.t.} \quad A_k\vec{p} + \vec{c}_k = 0$$

- For inequality, we can are the similar technique: $A_k\vec{p} + \vec{c}_k \geq 0$

The sequential quadratic programming

Algorithm: The sequential quadratic programming

1. Given $\vec{x}_0$.
2. For $k = 0, 1, 2, \ldots$ until converge
   1. Solve
   $$\min_{\vec{p}_k} \frac{1}{2}\vec{p}_k^T \nabla_{xx}^2 \mathcal{L} \vec{p}_k + \nabla f_k^T \vec{p}_k$$
   Subject to
   $$\nabla c_i(\vec{x}_k)^T \vec{p}_k + c_i(\vec{x}_k) = 0 \quad i \in \mathcal{E}$$
   $$\nabla c_i(\vec{x}_k)^T \vec{p}_k + c_i(\vec{x}_k) \geq 0 \quad i \in \mathcal{I}$$
   2. Set $\vec{x}_{k+1} = \vec{x}_k + \vec{p}_k$

Example

$$\min_{x_1,x_2} f(x_1,x_2) = e^{3x_1+4x_2} \text{ s.t. } c(x_1,x_2) = x_1^2 + x_2^2 = 1$$

The optimal solution is at $\vec{x}^* = \begin{bmatrix} -.6 \\ -.8 \end{bmatrix}, \lambda^* = -2.5e^{-5} \approx -0.168$

$$\nabla f = \begin{bmatrix} 3 \\ 4 \end{bmatrix} e^{3x_1+4x_2}, \nabla c = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}.$$

$$\nabla^2 f = \begin{bmatrix} 9 & 12 \\ 12 & 16 \end{bmatrix} e^{3x_1+4x_2}, \nabla c = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

The Lagrangian is $L(\vec{x}, \lambda) = f(\vec{x}) - \lambda c(\vec{x})$.

$$\nabla L_x(\vec{x}, \lambda) = \begin{bmatrix} 3f(\vec{x}) - 2\lambda x_1 \\ 4f(\vec{x}) - 2\lambda x_2 \end{bmatrix}$$

$$\nabla^2 L_{xx}(\vec{x}, \lambda) = \begin{bmatrix} 9f(\vec{x}) - 2\lambda & 12f(\vec{x}) \\ 12f(\vec{x}) & 16f(\vec{x}) - 2\lambda \end{bmatrix}$$

Example — continue

Let the initial guess be $\vec{x}_0 = [-0.7 \;\; -0.7]^T, \lambda_0 = -0.01$.

$$f(\vec{x}_0) = 0.0074, c(\vec{x}_0) = -0.02$$

$$\nabla^2 L_{xx}(\vec{x}_0, \lambda_0) = \begin{bmatrix} 0.087 & 0.089 \\ 0.089 & 0.139 \end{bmatrix}.$$

Solving the system, $\begin{bmatrix} \nabla^2_{xx}\mathcal{L} & -A_0^T \\ A_0 & 0 \end{bmatrix} \begin{bmatrix} \vec{p}_0 \\ \lambda_0 \end{bmatrix} = \begin{bmatrix} -\nabla f_0 + A_0^T \lambda_0 \\ -c_0 \end{bmatrix}$

$$\vec{p}_0 = \begin{bmatrix} .14 \\ -.15 \end{bmatrix}, \lambda_0 = -0.14808, \vec{x}_1 = \vec{x}_0 + \vec{p}_0 = \begin{bmatrix} -.56 \\ -.85 \end{bmatrix}$$

Problem of SQP

- For inequality constraints, linearization may cause inconsistent problem. For example, consider the constraints

$$\begin{aligned} c_1 &: \quad x - 1 \leq 0 \\ c_2 &: \quad x^2 - 4 \geq 0 \end{aligned}$$

  The feasible region is $x \leq -2$.
- The linearization of c_1 and c_2 at $x = 1$ becomes

$$\begin{array}{l} \nabla c_1^T \vec{p} + c_1(x) \leq 0 \\ \nabla c_2^T \vec{p} + c_2(x) \geq 0 \end{array} \Rightarrow \begin{array}{l} p \leq 0 \\ 2p - 3 \geq 0 \end{array},$$

  for which feasible region is empty.