MATH 4211/6211 – Optimization

Algorithms for Constrained Optimization

Xiaojing Ye
Department of Mathematics & Statistics
Georgia State University

We know that the gradient method proceeds as

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)}$$

where $\boldsymbol{d}^{(k)}$ is a descent direction (often chosen as a function of $\boldsymbol{g}^{(k)}$).

However, $\boldsymbol{x}^{(k+1)}$ is not necessarily in the feasible set Ω.

Hence the *projected gradient* (PG) method proceeds as

$$\boldsymbol{x}^{(k+1)} = \Pi(\boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{d}^{(k)})$$

in order that $\boldsymbol{x}^{(k)} \in \Omega$ for all k. Here $\Pi(\boldsymbol{x})$ is the *projection of* $\boldsymbol{x}$ *onto* Ω.

Definition. The projection Π onto Ω is defined by

$$\Pi(\boldsymbol{z}) = \arg\min_{\boldsymbol{x}\in\Omega} \|\boldsymbol{x} - \boldsymbol{z}\|$$

Namely, $\Pi(\boldsymbol{x})$ is the "closest point" in Ω to $\boldsymbol{x}$.

Note that $\Pi(\boldsymbol{x})$ is itself an optimization problem, which may not have closed-form or be easy to solve in most cases.

Example. Find the projection operators $\Pi(\boldsymbol{x})$ for the following sets Ω:

1. $\Omega = \{x \in \mathbb{R}^n : \|\boldsymbol{x}\|_\infty \leq 1\}$
2. $\Omega = \{x \in \mathbb{R}^n : a_i \leq x_i \leq b_i,\ \forall\, i\}$
3. $\Omega = \{x \in \mathbb{R}^n : \|\boldsymbol{x}\| \leq 1\}$
4. $\Omega = \{x \in \mathbb{R}^n : \|\boldsymbol{x}\| = 1\}$
5. $\Omega = \{x \in \mathbb{R}^n : \|\boldsymbol{x}\|_1 \leq 1\}$
6. $\Omega = \{x \in \mathbb{R}^n : \boldsymbol{Ax} = \boldsymbol{0}\}$ where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with $m \leq n$ is full rank.

Example. Consider the constrained optimization problem:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} \\ \text{subject to} \quad & \|\boldsymbol{x}\|^2 = 1 \end{aligned}$$

where $\boldsymbol{Q} \succ \mathbf{0}$. Apply the PG method with a fixed step size $\alpha > 0$ to this problem. Specifically:

- Write down the explicit formula of $\boldsymbol{x}^{(k+1)}$ in terms of $\boldsymbol{x}^{(k)}$ (assume never projecting $\mathbf{0}$).
- Is it possible to ensure convergence when α is sufficiently small?
- Show that if $\alpha \in (0, \frac{1}{\lambda_{\text{max}}})$ and $\boldsymbol{x}^{(0)}$ is not orthogonal to the smallest eigenvector corresponding to λ_{min}, then $\boldsymbol{x}^{(k)}$ converges. Here λ_{max} (λ_{min}) is the largest (smallest) eigenvalue of $\boldsymbol{Q}$.

Solution. We can see that the solution should be a unit eigenvector corresponding to λ_{min}.

Recall that $\Pi(\boldsymbol{x}) = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ for all $\boldsymbol{x} \neq \boldsymbol{0}$.

We also know $\nabla f(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x}$, and $\boldsymbol{x}^{(k)} - \alpha\nabla f(\boldsymbol{x}^{(k)}) = (\boldsymbol{I} - \alpha\boldsymbol{Q})\boldsymbol{x}^{(k)}$.

Therefore, PG with step size α is given by

$$\boldsymbol{x}^{(k+1)} = \beta_k(\boldsymbol{I} - \alpha\boldsymbol{Q})\boldsymbol{x}^{(k)}, \quad \text{where } \beta_k = \frac{1}{\|(\boldsymbol{I} - \alpha\boldsymbol{Q})\boldsymbol{x}^{(k)}\|}$$

Note that, if $\boldsymbol{x}^{(0)}$ is an eigenvector of $\boldsymbol{Q}$ corresponding to eigenvalue λ, then

$$\boldsymbol{x}^{(1)} = \beta_0(\boldsymbol{I} - \alpha\boldsymbol{Q})\boldsymbol{x}^{(0)} = \beta_0(1 - \alpha\lambda)\boldsymbol{x}^{(0)} = \boldsymbol{x}^{(0)}$$

and hence $\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(0)}$ for all k.

Solution (cont.) Denote $\lambda_1 \le \cdots \le \lambda_n$ the eigenvalues of $\boldsymbol{Q}$, and $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ the corresponding eigenvectors.

Now assume that

$$\boldsymbol{x}^{(k)} = y_1^{(k)}\boldsymbol{v}_1 + \cdots + y_n^{(k)}\boldsymbol{v}_n$$

Then we have

$$\boldsymbol{x}^{(k+1)} = \Pi((\boldsymbol{I}-\alpha\boldsymbol{Q})\boldsymbol{x}^{(k)}) = \beta_k y_1^{(k)}(1-\alpha\lambda_1)\boldsymbol{v}_1 + \cdots + \beta_k y_n^{(k)}(1-\alpha\lambda_n)\boldsymbol{v}_n$$

Denote $\beta^{(k)} = \prod_{j=0}^{k-1} \beta_j$, then

$$y_i^{(k)} = \beta_{k-1} y_i^{(k-1)}(1-\alpha\lambda_i) = \cdots = \beta^{(k)} y_i^{(0)}(1-\alpha\lambda_i)^k$$

Solution (cont.) Therefore, we have

$$\boldsymbol{x}^{(k)} = \sum_{i=1}^{n} y_i^{(k)} \boldsymbol{v}_i = y_1^{(k)} \left(\boldsymbol{v}_1 + \sum_{i=2}^{n} \frac{y_i^{(k)}}{y_1^{(k)}} \boldsymbol{v}_i \right)$$

Furthermore,

$$\frac{y_i^{(k)}}{y_1^{(k)}} = \frac{\beta^{(k)} y_i^{(0)} (1 - \alpha\lambda_i)^k}{\beta^{(k)} y_1^{(0)} (1 - \alpha\lambda_1)^k} = \frac{y_i^{(0)}}{y_1^{(0)}} \left(\frac{1 - \alpha\lambda_i}{1 - \alpha\lambda_1} \right)^k$$

Note that $y_1^{(0)} \neq 0$ (since $\boldsymbol{x}^{(0)}$ is not orthogonal to the eigenvector corresponding to λ_1). As $0 < \alpha < \frac{1}{\lambda_n}$, we have

$$0 < \frac{1 - \alpha\lambda_i}{1 - \alpha\lambda_1} < 1 \quad \Rightarrow \quad \left(\frac{1 - \alpha\lambda_i}{1 - \alpha\lambda_1} \right)^k \to 0 \text{ as } k \to \infty$$

for all $\lambda_i > \lambda_1$. Hence $\boldsymbol{x}^{(k)} \to \boldsymbol{v}_1$.

Projected gradient (PG) method for optimization with linear constraint:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

Then PG is given by

$$\boldsymbol{x}^{(k+1)} = \Pi(\boldsymbol{x}^{(k)} - \alpha_k \nabla f(\boldsymbol{x}^{(k)}))$$

where Π is the projection onto $\Omega := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$.

We first consider the *orthogonal projection* onto the hyperplane $\Psi = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$:

For any $\boldsymbol{v} \in \mathbb{R}^n$, the projection onto Ψ is the solution to

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{v}\|^2 \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \end{aligned}$$

Let $\boldsymbol{P} : \mathbb{R}^n \to \mathbb{R}^n$ denote this projector, i.e., $\boldsymbol{P}\boldsymbol{v}$ is the point on Ψ closest to $\boldsymbol{v}$.

The Lagrange function is

$$l(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{v}\|^2 + \boldsymbol{\lambda}^\top \boldsymbol{A}\boldsymbol{x}$$

Hence the Lagrange (KKT) condition is

$$\begin{aligned} (\boldsymbol{x} - \boldsymbol{v}) + \boldsymbol{A}^\top \boldsymbol{\lambda} &= \boldsymbol{0} \\ \boldsymbol{A}\boldsymbol{x} &= \boldsymbol{0} \end{aligned}$$

Left-multiplying the first equation by $\boldsymbol{A}$ and using $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, we obtain

$$\begin{aligned} \boldsymbol{\lambda} &= (\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{A}\boldsymbol{v} \\ \boldsymbol{x} &= (\boldsymbol{I} - \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{A})\boldsymbol{v} \end{aligned}$$

Denote the projector onto Ψ by

$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{A}$$

Thus, the projection of $\boldsymbol{v}$ onto Ψ is $\boldsymbol{P}\boldsymbol{v}$.

Proposition. The projector $\boldsymbol{P}$ has the following properties:

1. $\boldsymbol{P} = \boldsymbol{P}^\top$
2. $\boldsymbol{P}^2 = \boldsymbol{P}$.
3. $\boldsymbol{P}\boldsymbol{v} = \boldsymbol{0}$ iff $\exists\, \boldsymbol{\lambda} \in \mathbb{R}^m$ s.t. $\boldsymbol{v} = \boldsymbol{A}^\top \boldsymbol{\lambda}$. Namely $\mathcal{N}(\boldsymbol{P}) = \mathcal{R}(\boldsymbol{A}^\top)$.

Proof. Items 1 and 2 are easy to verify.

For item 3: ($\Rightarrow$) If $\boldsymbol{P}\boldsymbol{v} = \boldsymbol{0}$, then $\boldsymbol{v} = \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{A}\boldsymbol{v}$. Letting $\boldsymbol{\lambda} = (\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{A}\boldsymbol{v}$ yields $\boldsymbol{v} = \boldsymbol{A}^\top\boldsymbol{\lambda}$.

($\Leftarrow$) Suppose $\boldsymbol{v} = \boldsymbol{A}^\top\boldsymbol{\lambda}$, then

$$\boldsymbol{P}\boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{A})\boldsymbol{A}^\top\boldsymbol{\lambda} = \boldsymbol{A}^\top\boldsymbol{\lambda} - \boldsymbol{A}^\top\boldsymbol{\lambda} = \boldsymbol{0}.$$

Similar to the derivation of $\boldsymbol{P}$, we can obtain the projection onto Ω:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{v}\|^2 \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

(Write down the Lagrange function and KKT condition, and solve for $(\boldsymbol{x}, \boldsymbol{\lambda})$.)

The projection Π of $\boldsymbol{v}$ onto Ω is

$$\Pi(\boldsymbol{v}) = \boldsymbol{P}\boldsymbol{v} - \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{b}$$

Proposition. Let $\boldsymbol{x}^* \in \mathbb{R}^n$ be feasible (i.e., $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$), then $\boldsymbol{P}\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$ iff $\boldsymbol{x}^*$ satisfies the Lagrange condition.

Proof. We have

$$
\begin{aligned}
\boldsymbol{P}\nabla f(\boldsymbol{x}^*) = \boldsymbol{0} \quad &\Longleftrightarrow \quad \nabla f(\boldsymbol{x}^*) \in \mathcal{N}(\boldsymbol{P}) \\
&\Longleftrightarrow \quad \nabla f(\boldsymbol{x}^*) \in \mathcal{R}(\boldsymbol{A}^\top) \\
&\Longleftrightarrow \quad \nabla f(\boldsymbol{x}^*) = -\boldsymbol{A}^\top \boldsymbol{\lambda}^* \text{ for some } \boldsymbol{\lambda}^* \in \mathbb{R}^m
\end{aligned}
$$

Now we are ready to write down explicitly the PG:

$$
\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \Pi(\boldsymbol{x}^{(k)} - \alpha_k \nabla f(\boldsymbol{x}^{(k)})) && (\because \text{PG definition}) \\
&= \boldsymbol{P}(\boldsymbol{x}^{(k)} - \alpha_k \nabla f(\boldsymbol{x}^{(k)})) - \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{b} && (\because \text{Relation of } \Pi \text{ and } \boldsymbol{P}) \\
&= \boldsymbol{P}\boldsymbol{x}^{(k)} - \boldsymbol{A}^\top (\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{b} - \boldsymbol{P}\alpha_k \nabla f(\boldsymbol{x}^{(k)}) \\
&= \Pi(\boldsymbol{x}^{(k)}) - \alpha_k \boldsymbol{P} \nabla f(\boldsymbol{x}^{(k)}) && (\because \text{Relation of } \Pi \text{ and } \boldsymbol{P}) \\
&= \boldsymbol{x}^{(k)} - \alpha_k \boldsymbol{P} \nabla f(\boldsymbol{x}^{(k)}) && (\because \boldsymbol{x}^{(k)} \in \Omega)
\end{aligned}
$$

The only difference from standard gradient method is the additional $\boldsymbol{P}$.

Note that if $\boldsymbol{x}^{(0)} \in \Omega$, then $\boldsymbol{x}^{(k)} \in \Omega$ for all k.

Now we can consider the choice of α_k. For example, we can use the projected steepest descent (PSD) method:

$$\alpha_k = \operatorname*{arg\,min}_{\alpha > 0} f(\boldsymbol{x}^{(k)} - \alpha \boldsymbol{P} \nabla f(\boldsymbol{x}^{(k)}))$$

Theorem. Let $\boldsymbol{x}^{(k)}$ be generated by PSD. If $\boldsymbol{P}\nabla f(\boldsymbol{x}^{(k)}) \neq \mathbf{0}$, then $f(\boldsymbol{x}^{(k+1)}) < f(\boldsymbol{x}^{(k)})$.

Proof. For such $\boldsymbol{x}^{(k)}$, consider the line search function

$$\phi(\alpha) := f(\boldsymbol{x}^{(k)} - \alpha \boldsymbol{P}\nabla f(\boldsymbol{x}^{(k)})).$$

Then we have

$$\phi'(\alpha) = -\nabla f(\boldsymbol{x}^{(k)} - \alpha \boldsymbol{P}\nabla f(\boldsymbol{x}^{(k)}))^\top \boldsymbol{P}\nabla f(\boldsymbol{x}^{(k)}).$$

Hence

$$\begin{aligned}
\phi'(0) &= -\nabla f(\boldsymbol{x}^{(k)})^\top \boldsymbol{P}\nabla f(\boldsymbol{x}^{(k)}) \\
&= -\nabla f(\boldsymbol{x}^{(k)})^\top \boldsymbol{P}^2\nabla f(\boldsymbol{x}^{(k)}) \\
&= -\|\boldsymbol{P}\nabla f(\boldsymbol{x}^{(k)})\|^2 < 0,
\end{aligned}$$

and therefore $\phi(\alpha_k) < \phi(0)$, i.e., $f(\boldsymbol{x}^{(k+1)}) < f(\boldsymbol{x}^{(k)})$.

$\boldsymbol{P}\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ is sufficient for global optimality if f is convex:

Theorem. Let f be convex and $\boldsymbol{x}^*$ be feasible. Then $\boldsymbol{P}\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ iff $\boldsymbol{x}^*$ is a global minimizer.

Proof. From the previous proposition and convexity of f, we know

$$\begin{aligned} \boldsymbol{P}\nabla f(\boldsymbol{x}^*) = \mathbf{0} \quad &\Longleftrightarrow \quad \boldsymbol{x}^* \text{ satisfies the Lagrange condition} \\ &\Longleftrightarrow \quad \boldsymbol{x}^* \text{ is a global minimizer} \end{aligned}$$

Lagrange algorithm

We first consider the Lagrange algorithm for equality-constrained optimization:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \end{aligned}$$

where $f, \boldsymbol{h} \in \mathcal{C}^2$.

Recall the Lagrange function $l : \mathbb{R}^{n+m} \to \mathbb{R}$ is

$$l(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \boldsymbol{h}(\boldsymbol{x})^\top \boldsymbol{\lambda}.$$

We denote its Hessian with respect to $\boldsymbol{x}$ by

$$\nabla_{\boldsymbol{x}}^2 l(\boldsymbol{x}, \boldsymbol{\lambda}) = \nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x}) + D_{\boldsymbol{x}}^2 \boldsymbol{h}(\boldsymbol{x})^\top \boldsymbol{\lambda} \in \mathbb{R}^{n \times n}$$

Recall the Lagrange condition is

$$\begin{aligned} \nabla f(\boldsymbol{x}) + D\boldsymbol{h}(\boldsymbol{x})^\top \boldsymbol{\lambda} = \boldsymbol{0} \in \mathbb{R}^n \\ \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \in \mathbb{R}^m \end{aligned}$$

The **Lagrange algorithm** is given by

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} - \alpha_k (\nabla f(\boldsymbol{x}^{(k)}) + D\boldsymbol{h}(\boldsymbol{x}^{(k)})^\top \boldsymbol{\lambda}^{(k)}) \\ \boldsymbol{\lambda}^{(k+1)} &= \boldsymbol{\lambda}^{(k)} + \beta_k \boldsymbol{h}(\boldsymbol{x}^{(k)}) \end{aligned}$$

which is like “gradient descent for $\boldsymbol{x}$” and “gradient ascent for $\boldsymbol{\lambda}$” of l.

Here $\alpha_k, \beta_k \geq 0$ are step sizes. WLOG, we can assume $\alpha_k = \beta_k$ for all k by scaling $\boldsymbol{\lambda}^{(k)}$ properly.

It is easy to verify that, if $(\boldsymbol{x}^{(k)}, \boldsymbol{\lambda}^{(k)}) \to (\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$, then $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ satisfies the Lagrange condition.

We denote $\boldsymbol{w} = [\boldsymbol{x}; \boldsymbol{\lambda}] \in \mathbb{R}^{n+m}$ and

$$\boldsymbol{u}(\boldsymbol{w}) = \begin{bmatrix} \boldsymbol{x} - \alpha(\nabla f(\boldsymbol{x}) + D\boldsymbol{h}(\boldsymbol{x})^\top \boldsymbol{\lambda}) \\ \boldsymbol{\lambda} + \alpha \boldsymbol{h}(\boldsymbol{x}) \end{bmatrix} \in \mathbb{R}^{n+m}$$

Hence the Jacobian of $\boldsymbol{u}$ is

$$\nabla \boldsymbol{u}(\boldsymbol{w}) = \boldsymbol{I} + \alpha \begin{bmatrix} -\nabla_{\boldsymbol{x}}^2 l(\boldsymbol{x}, \boldsymbol{\lambda}) & -D\boldsymbol{h}(\boldsymbol{x})^\top \\ D\boldsymbol{h}(\boldsymbol{x}) & \boldsymbol{0} \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}$$

Note that

$$\boldsymbol{w}^* = [\boldsymbol{x}^*; \boldsymbol{\lambda}^*] \text{ is a KKT point} \quad \Longleftrightarrow \quad \boldsymbol{w}^* = \boldsymbol{u}(\boldsymbol{w}^*)$$

We denote

$$\boldsymbol{M} := \begin{bmatrix} -\nabla_{\boldsymbol{x}}^2 l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) & -D\boldsymbol{h}(\boldsymbol{x}^*)^\top \\ D\boldsymbol{h}(\boldsymbol{x}^*) & \boldsymbol{0} \end{bmatrix}$$

and hence $\nabla \boldsymbol{u}(\boldsymbol{w}^*) = \boldsymbol{I} + \alpha \boldsymbol{M}$.

Now we study the (local) convergence of the Lagrange algorithm when $\boldsymbol{x}^*$ is a regular point and $\nabla_{\boldsymbol{x}}^2 l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \succ \boldsymbol{0}$. For simplicity, we assume $\alpha_k = \alpha$ (constant step size).

Claim 1. $\|\nabla \boldsymbol{u}(\boldsymbol{w}^*)\| < 1$ if $\alpha > 0$ is sufficiently small.

Proof (Claim 1). It suffices to show real part of any eigenvalue of $\boldsymbol{M}$ is < 0.

Let λ be an eigenvalue of $\boldsymbol{M}$ and $\boldsymbol{w} = [\boldsymbol{x}; \boldsymbol{\lambda}] \in \mathbb{C}^{n+m}$ be a corresponding eigenvector, i.e., $\boldsymbol{M}\boldsymbol{w} = \lambda \boldsymbol{w}$. (Note $\boldsymbol{w} \neq \boldsymbol{0}$.)

If $\boldsymbol{x} = \boldsymbol{0}$, then

$$\boldsymbol{M}\boldsymbol{w} = \begin{bmatrix} -\nabla_{\boldsymbol{x}}^2 l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) & -D\boldsymbol{h}(\boldsymbol{x}^*)^\top \\ D\boldsymbol{h}(\boldsymbol{x}^*) & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -D\boldsymbol{h}(\boldsymbol{x}^*)^\top \boldsymbol{\lambda} \\ \boldsymbol{0} \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{\lambda} \end{bmatrix}$$

But $D\boldsymbol{h}(\boldsymbol{x}^*)$ has full row rank, so $\boldsymbol{\lambda} = \boldsymbol{0}$, and hence $\boldsymbol{w} = \boldsymbol{0}$, contradiction.

Proof (Claim 1) cont. Therefore $\boldsymbol{x} \neq \mathbf{0}$. We know *

$$\Re(\boldsymbol{w}^H \boldsymbol{M}\boldsymbol{w}) = \Re(\boldsymbol{w}^H \lambda \boldsymbol{w}) = \Re(\lambda)\|\boldsymbol{w}\|^2$$

On the other hand †

$$\Re(\boldsymbol{w}^H \boldsymbol{M}\boldsymbol{w}) = -\Re(\boldsymbol{x}^H \nabla_{\boldsymbol{x}}^2 l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)\boldsymbol{x}) < 0$$

Equating the two yields $\Re(\lambda) < 0$.

As all eigenvalues of $\boldsymbol{M}$ have negative real part, we know $\|\boldsymbol{I} + \alpha \boldsymbol{M}\| < 1$ for sufficiently small $\alpha > 0$.

This completes the proof of Claim 1.

* $\boldsymbol{w}^H$ is the complex conjugate of $\boldsymbol{w}$.

† Recall that if $\boldsymbol{Q} \succ \mathbf{0}$, then $\boldsymbol{x}^H \boldsymbol{Q}\boldsymbol{x} = \|\Re(\boldsymbol{x})\|_Q^2 + \|\Im(\boldsymbol{x})\|_Q^2$.

Claim 2. There exist $\eta > 0$ and $\kappa \in (0, 1)$ such that

$$\|\nabla \boldsymbol{u}(\boldsymbol{w})\| \le \kappa < 1, \quad \forall\, \boldsymbol{w} \in B(\boldsymbol{w}^*, \eta)$$

where $B(\boldsymbol{w}^*, \eta) = \{\boldsymbol{w} : \|\boldsymbol{w} - \boldsymbol{w}^*\| \le \eta\}$.

Proof (Claim 2). The claim follows $\|\nabla \boldsymbol{u}(\boldsymbol{w}^*)\| < 1$ in Claim 1 and the continuity of $\nabla \boldsymbol{u}$.

Claim 3. If $\boldsymbol{w}^{(0)} \in B(\boldsymbol{w}^*, \eta)$, then for all k there is

$$\|\boldsymbol{w}^{(k+1)} - \boldsymbol{w}^*\| \le \kappa \|\boldsymbol{w}^{(k)} - \boldsymbol{w}^*\|$$

Proof (Claim 3). Let $\boldsymbol{G} : \mathbb{R}^{n+m} \to \mathbb{R}^{(n+m)\times(n+m)}$ be the function s.t.

$$\boldsymbol{u}(\boldsymbol{w}^{(k)}) - \boldsymbol{u}(\boldsymbol{w}^*) = \boldsymbol{G}(\boldsymbol{w}^{(k)})(\boldsymbol{w}^{(k)} - \boldsymbol{w}^*)$$

from the Mean Value Theorem. Hence

$$\begin{aligned}
\|\boldsymbol{w}^{(k+1)} - \boldsymbol{w}^*\| &= \|\boldsymbol{u}(\boldsymbol{w}^{(k)}) - \boldsymbol{u}(\boldsymbol{w}^*)\| \\
&= \|\boldsymbol{G}(\boldsymbol{w}^{(k)})(\boldsymbol{w}^{(k)} - \boldsymbol{w}^*)\| \\
&\le \|\boldsymbol{G}(\boldsymbol{w}^{(k)})\| \cdot \|\boldsymbol{w}^{(k)} - \boldsymbol{w}^*\| \\
&\le \kappa \|\boldsymbol{w}^{(k)} - \boldsymbol{w}^*\|
\end{aligned}$$

Claim 3 implies that locally $\boldsymbol{w}^{(k)} \to \boldsymbol{w}^*$ at a linear rate.

Now consider **Lagrange algorithm** for inequality-constrained optimization:

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{g}(\boldsymbol{x}) \le \mathbf{0} \end{aligned}$$

The Lagrange function is

$$l(\boldsymbol{x}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})^\top \boldsymbol{\mu}$$

The Lagrange condition is

$$\begin{aligned} \nabla f(\boldsymbol{x}) + D\boldsymbol{g}(\boldsymbol{x})^\top \boldsymbol{\mu} &= \mathbf{0} \\ \boldsymbol{g}(\boldsymbol{x}) &\le \mathbf{0} \\ \boldsymbol{\mu} &\ge \mathbf{0} \\ \boldsymbol{g}(\boldsymbol{x})^\top \boldsymbol{\mu} &= 0 \end{aligned}$$

The Lagrange algorithm is given by

$$
\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} - \alpha_k(\nabla f(\boldsymbol{x}^{(k)}) + D\boldsymbol{g}(\boldsymbol{x}^{(k)})^\top \boldsymbol{\mu}^{(k)}) \\
\boldsymbol{\mu}^{(k+1)} &= [\boldsymbol{\mu}^{(k)} + \beta_k \boldsymbol{g}(\boldsymbol{x}^{(k)})]_+
\end{aligned}
$$

where $[\cdot]_+$ means $\mathsf{max}(\cdot, 0)$ componentwisely.

We denote $\boldsymbol{w} = [\boldsymbol{x}; \boldsymbol{\mu}] \in \mathbb{R}^{n+p}$ and

$$
\Pi(\boldsymbol{w}) = \begin{bmatrix} \boldsymbol{x} \\ [\boldsymbol{\mu}]_+ \end{bmatrix}, \qquad \boldsymbol{u}(\boldsymbol{w}) = \begin{bmatrix} \boldsymbol{x} - \alpha(\nabla f(\boldsymbol{x}) + D\boldsymbol{g}(\boldsymbol{x})^\top \boldsymbol{\mu}) \\ \boldsymbol{\mu} + \alpha \boldsymbol{g}(\boldsymbol{x}) \end{bmatrix}
$$

It is easy to verify that

$$
\boldsymbol{w}^* = [\boldsymbol{x}^*; \boldsymbol{\mu}^*] \text{ is a KKT point} \quad \Longleftrightarrow \quad \boldsymbol{w}^* = \Pi(\boldsymbol{u}(\boldsymbol{w}^*))
$$

Let $\boldsymbol{w}^*$ be a KKT point, and

$$\boldsymbol{g}(\boldsymbol{w}^*) = \begin{bmatrix} \boldsymbol{g}_A(\boldsymbol{w}^*) \\ \boldsymbol{g}_I(\boldsymbol{w}^*) \end{bmatrix} \in \mathbb{R}^p = \mathbb{R}^{p_1+p_2}, \quad \text{where} \quad \begin{array}{ll} \mathbf{0} = \boldsymbol{g}_A(\boldsymbol{w}) & \in \mathbb{R}^{p_1} \\ \mathbf{0} < \boldsymbol{g}_I(\boldsymbol{w}) & \in \mathbb{R}^{p_2} \end{array}$$

"A" and "I" stand for "active" and "inactive".

Similarly, denote

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_I \end{bmatrix}, \quad \boldsymbol{w}_A = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{\mu}_A \end{bmatrix}, \quad \boldsymbol{u}_A(\boldsymbol{w}_A) = \begin{bmatrix} \boldsymbol{x} - \alpha(\nabla f(\boldsymbol{x}) + D\boldsymbol{g}_A(\boldsymbol{x})^\top \boldsymbol{\mu}_A) \\ \mu_A + \alpha \boldsymbol{g}_A(\boldsymbol{x}) \end{bmatrix}$$

and hence

$$\nabla \boldsymbol{u}_A(\boldsymbol{w}_A) = \boldsymbol{I} + \alpha \begin{bmatrix} -\nabla^2_{\boldsymbol{x}} l(\boldsymbol{x}, \boldsymbol{\mu}_A) & -D\boldsymbol{g}_A(\boldsymbol{x})^\top \\ D\boldsymbol{g}_A(\boldsymbol{x}) & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(n+p_1)\times(n+p_1)}$$

Now we study the (local) convergence of the Lagrange algorithm when $\boldsymbol{x}^*$ is a regular point and $\nabla^2_{\boldsymbol{x}} l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \succ \boldsymbol{0}$. For simplicity, we assume $\alpha_k = \alpha$ (constant step size).

We again define $\boldsymbol{G}$ such that

$$\boldsymbol{u}(\boldsymbol{w}^{(k)}) - \boldsymbol{u}(\boldsymbol{w}^*) = \boldsymbol{G}(\boldsymbol{w}^{(k)})(\boldsymbol{w}^{(k)} - \boldsymbol{w}^*)$$

using Mean Value Theorem. Let

$$\boldsymbol{M} = \begin{bmatrix} -\nabla^2_{\boldsymbol{x}} l(\boldsymbol{x}^*, \boldsymbol{\mu}^*_A) & -D\boldsymbol{g}_A(\boldsymbol{x}^*)^\top \\ D\boldsymbol{g}_A(\boldsymbol{x}^*) & \boldsymbol{0} \end{bmatrix} \in \mathbb{R}^{(n+p_1)\times(n+p_1)}$$

Similar as before, we can show all eigenvalues of $\boldsymbol{M}$ have negative real part, and hence $\|\boldsymbol{I} + \alpha \boldsymbol{M}\| < 1$ for α sufficiently small.

Also note that $\boldsymbol{\mu}^*_I = \boldsymbol{0}$ as it corresponds to the inactive constraints.

Claim 1. There exist $\eta > 0$ and $\kappa_A \in (0, 1)$, such that

$$\begin{aligned} \|\nabla \boldsymbol{u}_A(\boldsymbol{w}_A)\| &\le \kappa_A \\ \boldsymbol{g}_I(\boldsymbol{x}) &\le -\delta \boldsymbol{e} \end{aligned}$$

for all $\boldsymbol{w} \in B(\boldsymbol{w}^*, \eta)$.

Proof. Note that $\boldsymbol{g}_I(\boldsymbol{w}^*) < \boldsymbol{0}$. Others are similar as before.

Now we set the following values:

- Let $\kappa = \max\{1, \|\boldsymbol{G}(\boldsymbol{w})\| : \boldsymbol{w} \in B(\boldsymbol{w}^*, \eta)\} \ge 1$
- Let $\varepsilon > 0$ be small enough such that $\varepsilon \kappa^{\varepsilon/(\alpha\delta)} \le \eta$.
- Let $k_0 = \lceil \varepsilon/(\alpha\delta) \rceil$.
- Let $\boldsymbol{w}^{(0)} \in B(\boldsymbol{w}^*, \varepsilon)$.

Claim 2. For any $k \le k_0$, there is $\|\boldsymbol{w}^{(k)} - \boldsymbol{w}^*\| \le \varepsilon\kappa^k$.

Proof (Claim 2). We use induction.

First, there is $\|\boldsymbol{w}^{(0)} - \boldsymbol{w}^*\| \le \varepsilon = \varepsilon\kappa^0$.

Assume the claim holds for k, then

$$\begin{aligned}
\|\boldsymbol{w}^{(k+1)} - \boldsymbol{w}^*\| &\le \|\boldsymbol{G}(\boldsymbol{w}^{(k)})\| \cdot \|\boldsymbol{w}^{(k)} - \boldsymbol{w}^*\| \\
&\le \kappa \cdot \|\boldsymbol{w}^{(k)} - \boldsymbol{w}^*\| \\
&\le \kappa \cdot (\varepsilon\kappa^k) \\
&= \varepsilon\kappa^{k+1}
\end{aligned}$$

which completes the proof of the claim.

From Claim 2, we know $\|\boldsymbol{w}^{(k)} - \boldsymbol{w}^*\| \le \eta$ for $k = 0, \dots, k_0$.

Claim 3. There is $\boldsymbol{\mu}_I^{(0)} \geq \cdots \geq \boldsymbol{\mu}_I^{(k_0)} = \mathbf{0}$.

Proof (Claim 3). We know $\boldsymbol{g}_I(\boldsymbol{x}^{(k)}) \leq -\delta \boldsymbol{e}$ for $k = 0, \dots, k_0$. Also

$$\boldsymbol{\mu}_I^{(k+1)} = [\boldsymbol{\mu}_I^{(k)} + \alpha \boldsymbol{g}_I(\boldsymbol{x}^{(k)})]_+ \leq [\boldsymbol{\mu}_I^{(k)} - \alpha\delta\boldsymbol{e}]_+ \leq \boldsymbol{\mu}_I^{(k)}$$

which implies that $\boldsymbol{\mu}_I^{(k)}$ is non-increasing.

Suppose $\mu_i^{(k_0)} > 0$ for some $i \in I$ (index set of inactive constraints), then

$$\begin{aligned} 0 < \mu_i^{(k_0)} &= \mu_i^{(k_0-1)} + \alpha g_i(\boldsymbol{x}^{(k_0-1)}) = \cdots \\ &= \mu_i^{(0)} + \alpha \sum_{k=0}^{k_0-1} g_i(\boldsymbol{x}^{(k)}) \leq \mu_i^{(0)} - \alpha\delta k_0 \leq \varepsilon - \alpha\delta k_0 \end{aligned}$$

since $\mu_i^{(0)} \leq \|\boldsymbol{w}^{(0)} - \boldsymbol{w}^*\| \leq \varepsilon$. But this contradicts to $k_0 = \lceil \frac{\epsilon}{\alpha\delta} \rceil \geq \frac{\epsilon}{\alpha\delta}$.

Therefore, within k_0 iterations, $\boldsymbol{\mu}_I^{(k)} = \mathbf{0}$.

Claim 4. For any $k \geq k_0$, there are

$$\boldsymbol{\mu}_I^{(k)} = \mathbf{0}$$
$$\|\boldsymbol{w}^{(k)} - \boldsymbol{w}^*\| \leq \eta$$
$$\|\boldsymbol{w}_A^{(k+1)} - \boldsymbol{w}_A^*\| \leq \kappa_A \|\boldsymbol{w}_A^{(k)} - \boldsymbol{w}_A^*\|$$

Proof (Claim 4). The first two hold for $k = k_0$ (by Claims 3 & 2 resp.), and

$$\begin{aligned}\|\boldsymbol{w}_A^{(k_0+1)} - \boldsymbol{w}_A^*\| &= \|\Pi(\boldsymbol{u}_A(\boldsymbol{w}_A^{(k_0)})) - \Pi(\boldsymbol{u}_A(\boldsymbol{w}_A^*))\| \\ &\leq \|\boldsymbol{u}_A(\boldsymbol{w}_A^{(k_0)}) - \boldsymbol{u}_A(\boldsymbol{w}_A^*)\| \\ &\leq \|\boldsymbol{G}_A(\boldsymbol{w}_A^{(k)})\| \cdot \|\boldsymbol{w}_A^{(k_0)} - \boldsymbol{w}_A^*\| \\ &\leq \kappa_A \cdot \|\boldsymbol{w}_A^{(k_0)} - \boldsymbol{w}_A^*\|\end{aligned}$$

Proof (Claim 4) cont.

Assume the results hold for $k \ge k_0$, then from $\boldsymbol{g}_I(\boldsymbol{w}^{(k)}) \le -\delta \boldsymbol{e}$, we have

$$\boldsymbol{\mu}_I^{(k+1)} = [\boldsymbol{\mu}_I^{(k)} + \alpha \boldsymbol{g}_I(\boldsymbol{x}^{(k)})]_+ \le [\mathbf{0} - \alpha\delta \boldsymbol{e}]_+ = \mathbf{0}$$

Note that this implies $\|\boldsymbol{w}_A^{(k+1)} - \boldsymbol{w}_A^*\| = \|\boldsymbol{w}^{(k+1)} - \boldsymbol{w}^*\|$ for all $k \ge k_0$.

Moreover, we have $\boldsymbol{w}_A^{(k+2)} = \Pi(\boldsymbol{u}_A(\boldsymbol{w}_A^{(k+1)}))$ and

$$\|\boldsymbol{w}_A^{(k+2)} - \boldsymbol{w}_A^*\| \le \kappa_A \cdot \|\boldsymbol{w}_A^{(k+1)} - \boldsymbol{w}_A^*\| \le \eta$$

which completes the proof.

Remark. Claim 4 implies that locally $\boldsymbol{w}^{(k)} \to \boldsymbol{w}^*$ at a linear rate: if $\boldsymbol{w}^{(0)}$ is sufficiently close to $\boldsymbol{w}^*$, then $\boldsymbol{w}^{(k)} \to \boldsymbol{w}^*$ linearly, provided that $\boldsymbol{x}^*$ is a regular KKT point and $\nabla_{\boldsymbol{x}}^2 l(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \succ \mathbf{0}$.

Penalty method

We consider constrained optimization

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in \Omega \end{aligned}$$

Note that such problem conceptually include optimization problems with equality and inequality constraints. For example, $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{g}(\boldsymbol{x}) \le \mathbf{0}\}$.

Instead of the constrained problem, we consider to impose penalty if $\boldsymbol{x} \in \Omega$ is violated:

$$\text{minimize} \;\; f(\boldsymbol{x}) + \gamma P(\boldsymbol{x})$$

where $P : \mathbb{R}^n \to \mathbb{R}_+$ is the penalty function, and $\gamma > 0$ is the penalty (weight) parameter.

Definition. The function $P : \mathbb{R}^n \to \mathbb{R}_+$ is called a **penalty function** if

1. P is continuous.
2. $P(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$.
3. $P(\boldsymbol{x}) = 0$ iff $\boldsymbol{x} \in \Omega$.

Example. Let $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0} \in \mathbb{R}^p\}$, then we can choose

$$P(\boldsymbol{x}) = \sum_{i=1}^{p} [g_i(\boldsymbol{x})]_+$$

$$P(\boldsymbol{x}) = \sum_{i=1}^{p} ([g_i(\boldsymbol{x})]_+)^2$$

and so on.

Example. Let $\boldsymbol{g}(x) = [g_1(x); g_2(x)]$ where $g_1(x) = x - 2$ and $g_2(x) = -(x+1)^3$. Consider the constraint set

$$\Omega = \{x \in \mathbb{R} : g_1(x) \le 0,\ g_2(x) \le 0\}$$

Then we have

$$[g_1(x)]_+ = \max\{0, g_1(x)\} = \begin{cases} 0 & \text{if } x \le 2 \\ x - 2 & \text{otherwise} \end{cases}$$

$$[g_2(x)]_+ = \max\{0, g_2(x)\} = \begin{cases} 0 & \text{if } x \ge -1 \\ -(x+1)^3 & \text{otherwise} \end{cases}$$

We can set

$$P(x) = [g_1(x)]_+ + [g_2(x)]_+ = \begin{cases} x - 2 & \text{if } x > 2 \\ 0 & \text{if } -1 \le x \le 2 \\ -(x+1)^3 & \text{if } x < -1 \end{cases}$$

Example. Consider the problem below with $\boldsymbol{Q} \succ \mathbf{0}$:

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} \\ \text{subject to} \quad & \|\boldsymbol{x}\|^2 = 1 \end{aligned}$$

We can set the penalty function $P(x) = (\|\boldsymbol{x}\|^2 - 1)^2$ (which is differentiable), and consider

$$\text{minimize} \quad \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} + \gamma(\|\boldsymbol{x}\|^2 - 1)^2$$

For any fixed $\gamma > 0$, the FONC of its solution $\boldsymbol{x}_\gamma$ is

$$2\boldsymbol{Q}\boldsymbol{x}_\gamma + 4\gamma(\|\boldsymbol{x}_\gamma\|^2 - 1)\boldsymbol{x}_\gamma = \mathbf{0}$$

which yields

$$\boldsymbol{Q}\boldsymbol{x}_\gamma = 2\gamma(1 - \|\boldsymbol{x}_\gamma\|^2)\boldsymbol{x}_\gamma = \lambda_\gamma \boldsymbol{x}_\gamma$$

where $\lambda_\gamma := 2\gamma(1 - \|\boldsymbol{x}_\gamma\|^2)$ is a scalar. This means $\lambda_\gamma \in (0, \lambda_{\text{max}}]$ is an eigenvalue of $\boldsymbol{Q}$, and $\boldsymbol{x}_\gamma$ is a corresponding eigenvector. Note that

$$0 < 1 - \|\boldsymbol{x}_\gamma\|^2 \le \frac{\lambda_{\text{max}}}{2\gamma} = \mathcal{O}\left(\frac{1}{\gamma}\right).$$

We have converted constrained problem into unconstrained ones. Now define

$$q(\gamma_k, \boldsymbol{x}) = f(\boldsymbol{x}) + \gamma_k P(\boldsymbol{x})$$
$$\boldsymbol{x}^{(k)} = \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^n} \ q(\gamma_k, \boldsymbol{x})$$

for every $k \in \mathbb{N}$.

The idea is to let γ_k increase (hence greater penalty) and apply an unconstrained optimization method to solve for $\boldsymbol{x}^{(k)}$ for each k.

Then we hope that an accumulation point‡ of $\{\boldsymbol{x}^{(k)}\}$ is a KKT point $\boldsymbol{x}^*$.

‡ $\boldsymbol{x}^*$ is called an *accumulation point* (also called *limit point*) of $\{\boldsymbol{x}^{(k)}\}$ if there exists a subsequence of $\boldsymbol{x}^{(k)}$ that converges to $\boldsymbol{x}^*$.

Now let $\gamma_k > 0$ be increasing, we have a series of claims.

Claim 1. $q(\gamma_k, \boldsymbol{x}^{(k)}) \leq q(\gamma_{k+1}, \boldsymbol{x}^{(k+1)})$.

Proof (Claim 1). Since $\boldsymbol{x}^{(k)}$ is optimal to $q(\gamma_k, \boldsymbol{x})$, we know

$$q(\gamma_k, \boldsymbol{x}^{(k)}) \leq q(\gamma_k, \boldsymbol{x}^{(k+1)})$$

Furthermore, since $\gamma_k < \gamma_{k+1}$, we know

$$\begin{aligned} q(\gamma_k, \boldsymbol{x}^{(k+1)}) &= f(\boldsymbol{x}^{(k+1)}) + \gamma_k P(\boldsymbol{x}^{(k+1)}) \\ &\leq f(\boldsymbol{x}^{(k+1)}) + \gamma_{k+1} P(\boldsymbol{x}^{(k+1)}) \\ &\leq q(\gamma_{k+1}, \boldsymbol{x}^{(k+1)}) \end{aligned}$$

Combining the two verifies the claim.

Claim 2. $P(\boldsymbol{x}^{(k+1)}) \le P(\boldsymbol{x}^{(k)})$.

Proof (Claim 2). By the optimality of $\boldsymbol{x}^{(k)}$ and $\boldsymbol{x}^{(k+1)}$ for their own problems, we know

$$q(\gamma_k, \boldsymbol{x}^{(k)}) \le q(\gamma_k, \boldsymbol{x}^{(k+1)})$$
$$q(\gamma_{k+1}, \boldsymbol{x}^{(k+1)}) \le q(\gamma_{k+1}, \boldsymbol{x}^{(k)})$$

which are

$$f(\boldsymbol{x}^{(k)}) + \gamma_k P(\boldsymbol{x}^{(k)}) \le f(\boldsymbol{x}^{(k+1)}) + \gamma_k P(\boldsymbol{x}^{(k+1)})$$
$$f(\boldsymbol{x}^{(k+1)}) + \gamma_{k+1} P(\boldsymbol{x}^{(k+1)}) \le f(\boldsymbol{x}^{(k)}) + \gamma_{k+1} P(\boldsymbol{x}^{(k)})$$

Adding the two above yields

$$(\gamma_{k+1} - \gamma_k) P(\boldsymbol{x}^{(k+1)}) \le (\gamma_{k+1} - \gamma_k) P(\boldsymbol{x}^{(k)})$$

Recalling $\gamma_{k+1} - \gamma_k > 0$ completes the proof.

Claim 3. $f(\boldsymbol{x}^{(k+1)}) \geq f(\boldsymbol{x}^{(k)})$.

Proof (Claim 3). Since $q(\gamma_k, \boldsymbol{x}^{(k)}) \leq q(\gamma_k, \boldsymbol{x}^{(k+1)})$, we know

$$f(\boldsymbol{x}^{(k)}) + \gamma_k P(\boldsymbol{x}^{(k)}) \leq f(\boldsymbol{x}^{(k+1)}) + \gamma_k P(\boldsymbol{x}^{(k+1)})$$

From Claim 2, we know $P(\boldsymbol{x}^{(k+1)}) \leq P(\boldsymbol{x}^{(k)})$, hence

$$f(\boldsymbol{x}^{(k+1)}) \geq f(\boldsymbol{x}^{(k)}) + \gamma_k (P(\boldsymbol{x}^{(k)}) - P(\boldsymbol{x}^{(k+1)})) \geq f(\boldsymbol{x}^{(k)})$$

Claim 4. $f(\boldsymbol{x}^*) \geq q(\gamma_k, \boldsymbol{x}^{(k)}) \geq f(\boldsymbol{x}^{(k)})$.

Proof (Claim 4). We know $P(\boldsymbol{x}^*) = 0$, and hence

$$\begin{aligned} f(\boldsymbol{x}^*) &= q(\gamma_k, \boldsymbol{x}^*) \\ &\geq q(\gamma_k, \boldsymbol{x}^{(k)}) \\ &= f(\boldsymbol{x}^{(k)}) + \gamma_k P(\boldsymbol{x}^{(k)}) \\ &\geq f(\boldsymbol{x}^{(k)}) \end{aligned}$$

Theorem. Suppose f is continuous and $\gamma_k \uparrow \infty$. Then any accumulation point of $\{\boldsymbol{x}^{(k)}\}$ is a solution to the constrained problem.

Proof. For simplicity, let $\boldsymbol{x}^{(k)}$ denote the subsequence which converges to $\widehat{\boldsymbol{x}}$.

Since $f(\boldsymbol{x}^{(k)}) \le f(\boldsymbol{x}^*)$ for all k (by Claim 4), we know

$$f(\boldsymbol{x}^*) \ge \lim_{k\to\infty} f(\boldsymbol{x}^{(k)}) = f(\widehat{\boldsymbol{x}})$$

Note that $q(\gamma_k, \boldsymbol{x}^{(k)})$ is nondecreasing in k (by Claim 1) and bounded above by $f(\boldsymbol{x}^*)$ (by Claim 4), we know $q(\gamma_k, \boldsymbol{x}^{(k)}) \uparrow q^*$ for some $q^* \in \mathbb{R}$. Hence,

$$\gamma_k P(\boldsymbol{x}^{(k)}) = q(\gamma_k, \boldsymbol{x}^{(k)}) - f(\boldsymbol{x}^{(k)}) \to q^* - f(\widehat{\boldsymbol{x}})$$

Since $\gamma_k \to \infty$, we know $P(\boldsymbol{x}^{(k)}) \to 0$. Since P is continuous, we know $P(\widehat{\boldsymbol{x}}) = 0$, i.e., $\widehat{\boldsymbol{x}}$ is feasible. Therefore $\widehat{\boldsymbol{x}}$ is optimal since $f(\widehat{\boldsymbol{x}}) \le f(\boldsymbol{x}^*)$.

Penalty method requires solving one instance of

$$\text{minimize} \;\; f(\boldsymbol{x}) + \gamma P(\boldsymbol{x})$$

with $\gamma = \gamma_k$ for every k.

Is it possible to obtain the solution with a single γ?

Definition. We call P an **exact penalty** if there exists $\gamma > 0$ such that the solution $\boldsymbol{x}^*$ of the unconstrained problem

$$\text{minimize} \;\; f(\boldsymbol{x}) + \gamma P(\boldsymbol{x})$$

is also a solution of the constrained problem

$$\begin{aligned} \text{minimize} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in \Omega \end{aligned}$$

However it turns out that it may be necessary for an exact penalty P to be non-differentiable.

Proposition. Let Ω be convex, $\boldsymbol{x}^*$ is on the boundary of Ω. If there exists a feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$ such that $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) > 0$, then an exact penalty P must be non-differentiable.

Proof. Suppose not, then $\nabla P(\boldsymbol{x}^*) = 0$ since $P(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \Omega$. Let $g(\boldsymbol{x}) = f(\boldsymbol{x}) + \gamma P(\boldsymbol{x})$, then

$$\nabla g(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*) + \gamma \nabla P(\boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*)$$

and hence $\boldsymbol{d}^\top g(\boldsymbol{x}^*) = \boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) > 0$, which means $\boldsymbol{x}^*$ is not a local minimizer of g, contradiction.

Example. Consider the problem

$$\begin{aligned} \text{minimize} \quad & 5 - 3x \\ \text{subject to} \quad & x \in [0, 1] \end{aligned}$$

We can see $x^* = 1$ which is on the boundary, and $f'(x^*) = -3$ aligns with the feasible direction $d = -1$ at x^*.

If we use a differentiable penalty function P, then $P'(x^*) = 0$. Let

$$g(x) = f(x) + \gamma P(x),$$

then $g'(x^*) = f'(x^*) + \gamma P'(x^*) = -3 \neq 0$, which means P cannot be an exact penalty function.

Remark. However, if $\boldsymbol{d}^\top \nabla f(\boldsymbol{x}^*) \leq 0$ for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}$, we may still be able to find a differentiable exact penalty function P.