



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Robotic Systems II

Lecture 3: Constrained Optimization Methods

Konstantinos Chatzilygeroudis - costashatz@upatras.gr

Department of Electrical and Computer Engineering
University of Patras

Template made by Panagiotis Papagiannopoulos



■ Lectures:

- 10 Lectures
- 1 Recitation Lecture
- 1-2 Seminars
- 10 Lab Exercises

■ Examination:

- 4 Homeworks (**40% of total grade**)
- Oral Exam (**60% of total grade**)

■ Office Hours:

- **Tue & Thu (09:00-11:00)**
- 24/7 by email (costashatz@upatras.gr, subject: *ECE_RSII_AM*)

■ Material and Announcements



What is this course about?

We could name the course “**Applied Optimal Control**”:

- Optimization
- Control of Linear Systems (LQR)
- Model-Predictive Control
- Linearization of Non-Linear Systems
- Trajectory Optimization (Direct Collocation)
- Optimal Control on Manifolds
- Differential Dynamic Programming
- Locomotion, Contacts, Complex Robots
- **We focus on implementation/robotics applications. You are required to write code!** Not a theory course!

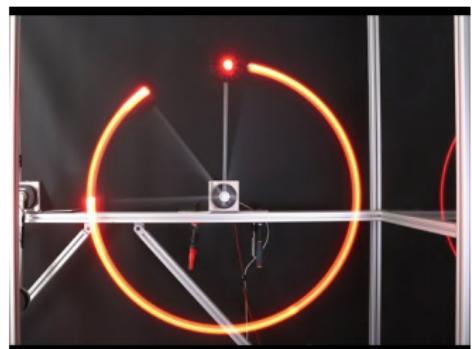
- **3-hour lectures:**

- A lot of theory! Hints for personal reading!
- Live code examples + analyses!
- Ask questions please!

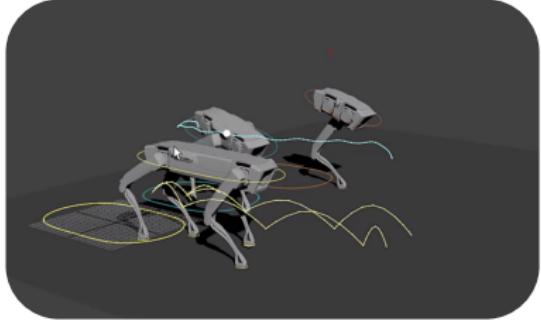
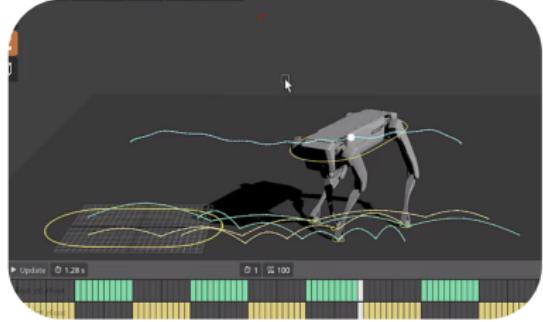
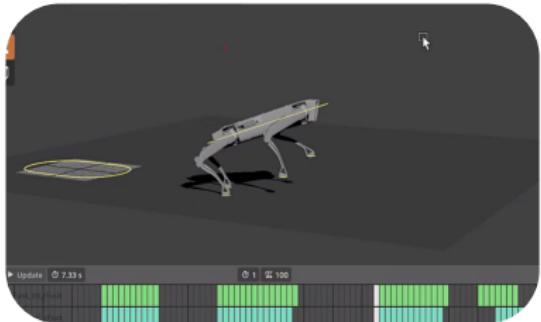
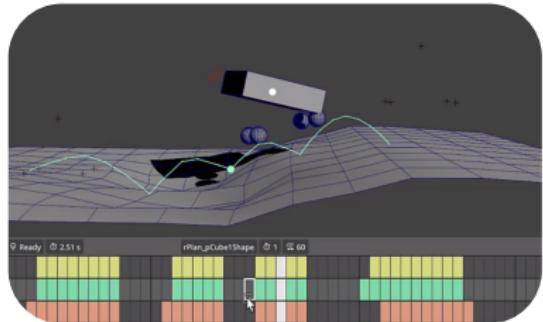
- **2-hour lab exercises:**

- Do not count for final grade! **You need to participate in 60%!**
- Small code examples that you need to fill
- You need to deliver code even if it doesn't run
- Ask questions, experiment!

Why study optimal control?



What will you be able to achieve?



Credits: Ragdoll Dynamics (Imbalance Ltd)



Duality

$$\begin{aligned} & \min_{\mathbf{x}} f(\mathbf{x}) \\ \text{s.t. } & h(\mathbf{x}) = 0 \end{aligned}$$

We could write the problem as:

$$\min_{\mathbf{x}} f(\mathbf{x}) + P_{\infty}(h(\mathbf{x}))$$

$$\text{where } P_{\infty}(h(\mathbf{x})) = \begin{cases} 0 & \text{for } h(\mathbf{x}) = 0 \\ \infty & \text{otherwise} \end{cases}.$$

Duality (2)

Of course, this is “useless” in practical applications, but the above is equivalent to:

$$\min_{\mathbf{x}} \max_{\boldsymbol{\lambda}} \underbrace{f(\mathbf{x}) + \boldsymbol{\lambda}^T h(\mathbf{x})}_{\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})}$$

Why?

Duality (2)

Of course, this is “useless” in practical applications, but the above is equivalent to:

$$\min_{\mathbf{x}} \max_{\boldsymbol{\lambda}} \underbrace{f(\mathbf{x}) + \boldsymbol{\lambda}^T h(\mathbf{x})}_{\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})}$$

Why?

KKT Conditions (equivalent to the above) define a saddle point in $(\mathbf{x}, \boldsymbol{\lambda})$: the KKT system should have N (dimension of \mathbf{x}) positive eigenvalues and M (dimension of $\boldsymbol{\lambda}$) negative eigenvalues at the optimum (“quasi-definite”).

Strong duality: We can swap the min and max. Only when the problem is **strictly convex**.

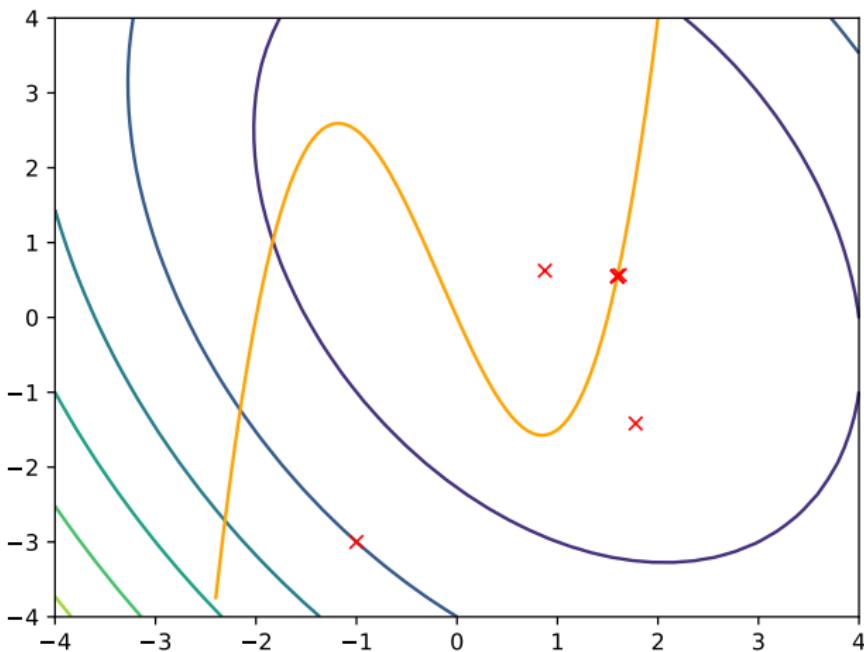
Regularization in Constrained Optimization

We now know how to regularize a constrained KKT system:

$$\begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \mathbf{x}^2} \Big|_{\mathbf{x}, \boldsymbol{\lambda}} + \beta \mathbf{I} & \left(\frac{\partial h}{\partial \mathbf{x}} \Big|_{\mathbf{x}} \right)^T \\ \frac{\partial h}{\partial \mathbf{x}} \Big|_{\mathbf{x}} & -\beta \mathbf{I} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \\ -h(\mathbf{x}) \end{bmatrix}$$

for $\beta > 0$. We apply this while the KKT system is not “quasi-definite!”

Regularized Newton's Method - Code Example



$$\begin{aligned} & \min_x f(x) \\ \text{s.t. } & g(x) \leq 0 \end{aligned}$$

where $f(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}$, $g(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}^M$. We define the **Lagrangian** as follows:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\mu}^T g(\mathbf{x})$$

First Order Necessary (KKT) Conditions:

- | | |
|---|---|
| 1) $\nabla_{\mathbf{x}} f(\mathbf{x}) + \left(\frac{\partial g}{\partial \mathbf{x}} \right)^T \boldsymbol{\mu} = 0,$
2) $g(\mathbf{x}) \leq 0,$
3) $\boldsymbol{\mu} \geq 0,$
4) $\boldsymbol{\mu}^T g(\mathbf{x}) = 0,$ | “stationarity”
“primal feasibility”
“dual feasibility”
“complementarity” |
|---|---|

Augmented Lagrangian Method

We define the **Augmented Lagrangian** as:

$$\mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\mu}^T g(\mathbf{x}) + \frac{\rho}{2} \|\max(0, g(\mathbf{x}))\|^2$$

$\frac{\rho}{2} \|\max(0, g(\mathbf{x}))\|^2$ is called the **penalty** term.

$$\begin{aligned}\nabla_{\mathbf{x}} \mathcal{L}_\rho &= \left(\frac{\partial f}{\partial \mathbf{x}} + \boldsymbol{\mu}^T \frac{\partial g}{\partial \mathbf{x}} + \rho g(\mathbf{x})^T \frac{\partial g}{\partial \mathbf{x}} \right)^T \\ &= \left(\frac{\partial f}{\partial \mathbf{x}} + \left(\boldsymbol{\mu} + \rho g(\mathbf{x}) \right)^T \frac{\partial g}{\partial \mathbf{x}} \right)^T = 0\end{aligned}$$

So we update $\boldsymbol{\mu}$ at each iteration as follows:

$$\boldsymbol{\mu} = \boldsymbol{\mu} + \rho g(\mathbf{x})$$

Augmented Lagrangian Method (2)

$$1) \quad \mathbf{x}_{k+1} = \min_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\mu}_k)$$

$$2) \quad \boldsymbol{\mu}_{k+1} = \max\left(\mathbf{0}, \boldsymbol{\mu}_k + \rho g(\mathbf{x}_{k+1})\right)$$

$$3) \quad \rho = \alpha \rho, \text{ with } \alpha \in \mathbb{R}^+$$

Usually $\alpha \approx 10$ and we use (2) only when the constraints are violated. The method comes with strong convergence guarantees.

How can we solve (1)?

Augmented Lagrangian Method - Minimizing \mathcal{L}_ρ

- We need to solve $\min_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\mu})$
- Important to note that $\boldsymbol{\mu}$ and ρ are constants!
- Let's apply Newton's Method:

$$\nabla_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}_k + \Delta \mathbf{x}, \boldsymbol{\mu}_k) \approx$$

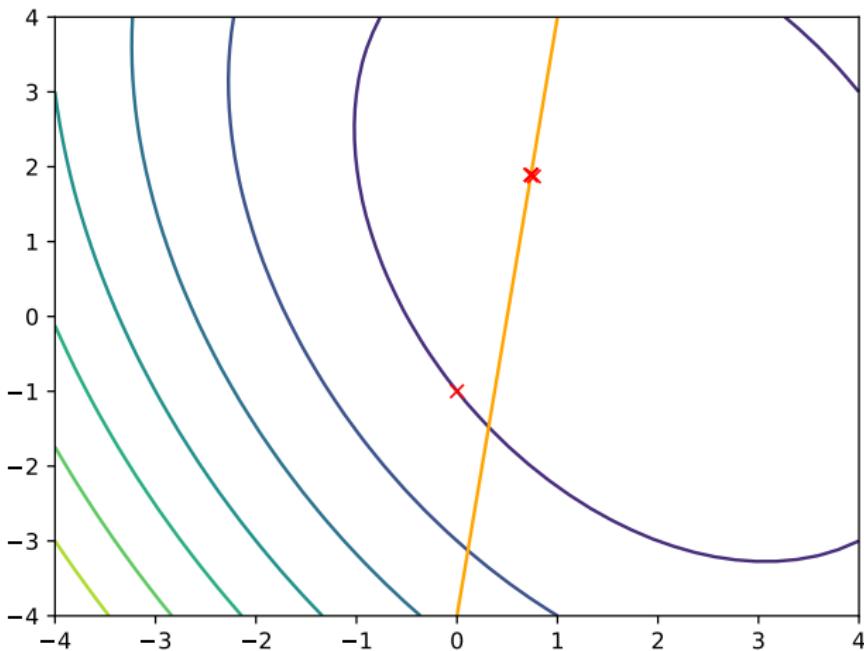
$$\nabla_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}_k, \boldsymbol{\mu}_k) + \frac{\partial}{\partial \mathbf{x}} \left(\nabla_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}_k, \boldsymbol{\mu}_k) \right) \Delta \mathbf{x} = 0$$

We can use the Gauss-Newton version and:

$$\frac{\partial}{\partial \mathbf{x}} \left(\nabla_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}_k, \boldsymbol{\mu}_k) \right) \approx \nabla_{\mathbf{x}\mathbf{x}}^2 f(\mathbf{x}_k) + \rho \left(\frac{\partial g}{\partial \mathbf{x}} \Big|_{\mathbf{x}_k} \right)^T \frac{\partial g}{\partial \mathbf{x}} \Big|_{\mathbf{x}_k}$$

$$\Delta \mathbf{x} = - \left(\nabla_{\mathbf{x}\mathbf{x}}^2 f(\mathbf{x}_k) + \rho \left(\frac{\partial g}{\partial \mathbf{x}} \Big|_{\mathbf{x}_k} \right)^T \frac{\partial g}{\partial \mathbf{x}} \Big|_{\mathbf{x}_k} \right)^{-1} \nabla_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}_k, \boldsymbol{\mu}_k)$$

Augmented Lagrangian Method - Code Example



Full constrained minimization:

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g(\mathbf{x}) \leq 0 \\ & h(\mathbf{x}) = 0 \end{aligned}$$

The **Lagrangian** is given:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T h(\mathbf{x}) + \boldsymbol{\mu}^T g(\mathbf{x})$$

First Order Necessary (KKT) Conditions:

- 1) $\nabla_{\mathbf{x}} f(\mathbf{x}) + \left(\frac{\partial h}{\partial \mathbf{x}} \right)^T \boldsymbol{\lambda} + \left(\frac{\partial g}{\partial \mathbf{x}} \right)^T \boldsymbol{\mu} = 0,$ “stationarity”
- 2) $g(\mathbf{x}) \leq 0$ and $h(\mathbf{x}) = 0,$ “primal feasibility”
- 3) $\boldsymbol{\mu} \geq 0,$ “dual feasibility”
- 4) $\boldsymbol{\mu}^T g(\mathbf{x}) = 0,$ “complementarity”

Merit Functions

- How can we apply line search in the constrained case?

- How can we apply line search in the constrained case?
- We define a **merit function**, $P(\mathbf{x})$
- Then we apply the Armijo rule to it

$$\alpha = 1$$

while $P(\mathbf{x}_k + \alpha \Delta \mathbf{x}) > P(\mathbf{x}_k) + b\alpha \nabla_{\mathbf{x}} P(\mathbf{x}_k)^T \Delta \mathbf{x}$:

$$\alpha = c\alpha$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha \Delta \mathbf{x}$$

Usual Merit Functions

- KKT residual

$$P(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2} \left\| \begin{bmatrix} \nabla_{\mathbf{x}} \mathcal{L} \\ h(\mathbf{x}) \\ \min(0, g(\mathbf{x})) \end{bmatrix} \right\|^2$$

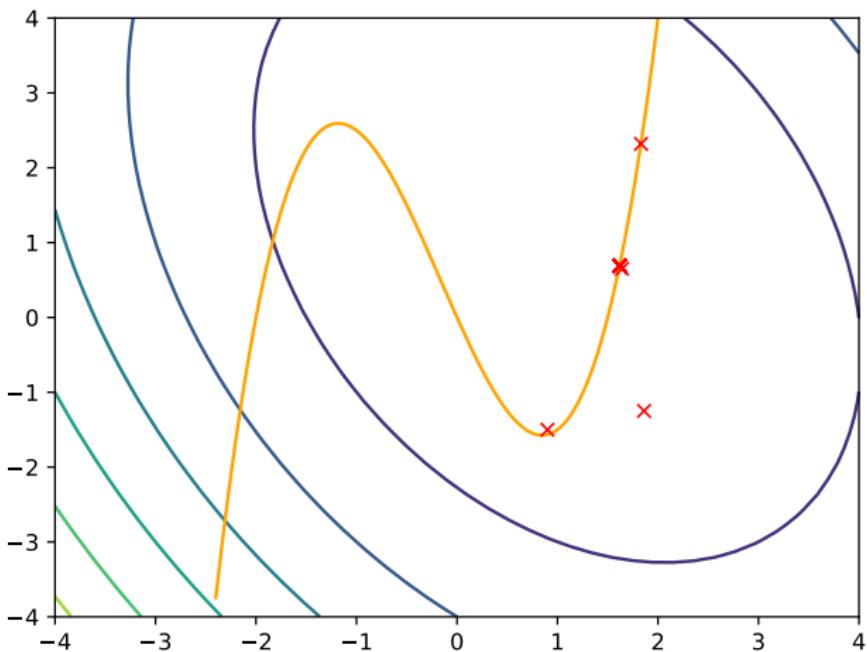
- Weighted Sum

$$P(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \frac{1}{2} w \left\| \begin{bmatrix} h(\mathbf{x}) \\ \min(0, g(\mathbf{x})) \end{bmatrix} \right\|^2$$

- Augmented Lagrangian

$$P(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathcal{L}_{\rho}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

Merit Functions - Code Example



ALM Pseudocode for Constrained Minimization

Input: Initial \mathbf{x}_0 , $\boldsymbol{\lambda}_0$ (equality), $\boldsymbol{\mu}_0 \geq \mathbf{0}$ (inequality); penalties $\rho_h, \rho_g > 0$; growth factors $\alpha_h, \alpha_g > 1$; tolerances $\varepsilon_{\text{stat}}, \varepsilon_{\text{fe},h}, \varepsilon_{\text{fe},g}$; max outer K

Output: Approximate solution \mathbf{x}_k to $\min f(\mathbf{x})$ s.t. $\mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}$

```
1 for  $k = 0, 1, \dots, K - 1$  do
2     // Define the augmented Lagrangian (your notation)
3      $\mathcal{L}_{\rho_h, \rho_g}(\mathbf{x}, \boldsymbol{\lambda}_k, \boldsymbol{\mu}_k) = f(\mathbf{x}) + \boldsymbol{\lambda}_k^\top \mathbf{h}(\mathbf{x}) + \frac{\rho_h}{2} \|\mathbf{h}(\mathbf{x})\|^2 + \boldsymbol{\mu}_k^\top \mathbf{g}(\mathbf{x}) + \frac{\rho_g}{2} \|[\mathbf{g}(\mathbf{x})]_+\|^2$ 
4     // (1) Primal update: approximately minimize w.r.t.  $\mathbf{x}$ 
5     Solve (approximately)  $\mathbf{x}_{k+1} \approx \arg \min_{\mathbf{x}} \mathcal{L}_{\rho_h, \rho_g}(\mathbf{x}, \boldsymbol{\lambda}_k, \boldsymbol{\mu}_k)$ 
6     // Evaluate constraint residuals at  $\mathbf{x}_{k+1}$ 
7      $\mathbf{h}_{k+1} \leftarrow \mathbf{h}(\mathbf{x}_{k+1}), \quad \mathbf{g}_{k+1} \leftarrow \mathbf{g}(\mathbf{x}_{k+1})$ 
8     // (2) Dual updates (project  $\boldsymbol{\mu}$  onto  $\mathbb{R}_+^m$ )
9      $\boldsymbol{\lambda}_{k+1} \leftarrow \boldsymbol{\lambda}_k + \rho_h \mathbf{h}_{k+1}$ 
10     $\boldsymbol{\mu}_{k+1} \leftarrow \max(0, \boldsymbol{\mu}_k + \rho_g \mathbf{g}_{k+1})$ 
11    // (3) Penalty updates (e.g. simple multiplicative growth)
12     $\rho_h \leftarrow \alpha_h \rho_h; \quad \rho_g \leftarrow \alpha_g \rho_g$ 
13    // Convergence (feasibility + (approx.) stationarity)
14    if  $\|\mathbf{h}_{k+1}\| \leq \varepsilon_{\text{fe},h}$  and  $\|[\mathbf{g}_{k+1}]_+\| \leq \varepsilon_{\text{fe},g}$  and  $\|\nabla f(\mathbf{x}_{k+1}) + (\frac{\partial \mathbf{h}}{\partial \mathbf{x}})^\top \boldsymbol{\lambda}_{k+1} + (\frac{\partial \mathbf{g}}{\partial \mathbf{x}})^\top \boldsymbol{\mu}_{k+1}\| \leq \varepsilon_{\text{stat}}$ 
15        then
16            break
17
18 return  $\mathbf{x}_k$ 
```

$$\begin{aligned} & \min_{\mathbf{x}} f(\mathbf{x}) \\ \text{s.t. } & \mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{0} \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\mathbf{b} \in \mathbb{R}^M$ and $f : \mathbb{R}^N \rightarrow \mathbb{R}$ is convex.

How do we solve this?

$$\begin{aligned} & \min_{\mathbf{x}} f(\mathbf{x}) \\ \text{s.t. } & \mathbf{A}\mathbf{x} - \mathbf{b} = \mathbf{0} \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^N$, $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\mathbf{b} \in \mathbb{R}^M$ and $f : \mathbb{R}^N \rightarrow \mathbb{R}$ is convex.

How do we solve this?

We can use the **Augmented Lagrangian**:

$$\mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T (\mathbf{A}\mathbf{x} - \mathbf{b}) + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$$

Can we do better?

It turns out we can if we can split the problem into subparts:

$$\min_{\mathbf{y}, \mathbf{z}} f(\mathbf{y}) + g(\mathbf{z})$$

$$\text{s.t. } \mathbf{A}\mathbf{y} + \mathbf{B}\mathbf{z} - \mathbf{b} = \mathbf{0}$$

where $\mathbf{y} \in \mathbb{R}^Q$, $\mathbf{z} \in \mathbb{R}^P$, $\mathbf{A} \in \mathbb{R}^{M \times Q}$, $\mathbf{B} \in \mathbb{R}^{M \times P}$, $\mathbf{b} \in \mathbb{R}^M$,
 $f : \mathbb{R}^Q \rightarrow \mathbb{R}$ is convex and $g : \mathbb{R}^P \rightarrow \mathbb{R}$ is convex. This is possible
in many ways, for example one usual case is:

$$\min_{\mathbf{y}, \mathbf{z}} f(\mathbf{y}) + g(\mathbf{z})$$

$$\text{s.t. } \mathbf{y} = \mathbf{z}$$

Alternating Direction Method of Multipliers:

- 1) $\mathbf{y}_{k+1} = \min_{\mathbf{y}} \mathcal{L}_\rho(\mathbf{y}, \mathbf{z}_k, \boldsymbol{\lambda}_k)$
- 2) $\mathbf{z}_{k+1} = \min_{\mathbf{z}} \mathcal{L}_\rho(\mathbf{y}_{k+1}, \mathbf{z}, \boldsymbol{\lambda}_k)$
- 3) $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \rho(\mathbf{A}\mathbf{y}_{k+1} + \mathbf{B}\mathbf{z}_{k+1} - \mathbf{b})$
- 4) $\rho = \alpha\rho$, with $\alpha \in \mathbb{R}^+$

What is the difference with solving directly with Augmented Lagrangian Method?

Quadratic Programming (QP)

$$\begin{aligned}\min_{\mathbf{x}} f(\mathbf{x}) &= \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{q}^T \mathbf{x} \\ \text{s.t. } \mathbf{A} \mathbf{x} - \mathbf{b} &= \mathbf{0} \\ \mathbf{C} \mathbf{x} - \mathbf{d} &\leq \mathbf{0}\end{aligned}$$

where $\mathbf{x}, \mathbf{q} \in \mathbb{R}^N$, $\mathbf{Q} > 0 \in \mathbb{R}^{N \times N}$. Let's solve this with the **Augmented Lagrangian Method**:

Quadratic Programming (QP)

$$\begin{aligned}\min_{\mathbf{x}} f(\mathbf{x}) &= \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{q}^T \mathbf{x} \\ \text{s.t. } \mathbf{A} \mathbf{x} - \mathbf{b} &= \mathbf{0} \\ \mathbf{C} \mathbf{x} - \mathbf{d} &\leq \mathbf{0}\end{aligned}$$

where $\mathbf{x}, \mathbf{q} \in \mathbb{R}^N$, $\mathbf{Q} > 0 \in \mathbb{R}^{N \times N}$. Let's solve this with the **Augmented Lagrangian Method**:

$$\begin{aligned}\mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= f(\mathbf{x}) + \boldsymbol{\lambda}^T (\mathbf{A} \mathbf{x} - \mathbf{b}) + \boldsymbol{\mu}^T (\mathbf{C} \mathbf{x} - \mathbf{d}) \\ &\quad + \frac{\rho}{2} \|\mathbf{A} \mathbf{x} - \mathbf{b}\|^2 + \frac{\rho}{2} \|\max(\mathbf{0}, \mathbf{C} \mathbf{x} - \mathbf{d})\|^2\end{aligned}$$

KKT Conditions:

- 1) $\mathbf{Q} \mathbf{x} + \mathbf{q} + \mathbf{A}^T \boldsymbol{\lambda} + \mathbf{C}^T \boldsymbol{\mu} = \mathbf{0}$
- 2) $\mathbf{A} \mathbf{x} - \mathbf{b} = \mathbf{0}$ and $\mathbf{C} \mathbf{x} - \mathbf{d} \leq \mathbf{0}$
- 3) $\boldsymbol{\mu} \geq \mathbf{0}$
- 4) $\boldsymbol{\mu}^T (\mathbf{C} \mathbf{x} - \mathbf{d}) = \mathbf{0}$

What other methods are there?

- **Active Set Methods:**

- Try to guess which constraints are active (equal to zero) or not
- Solve equality constrained optimization only with active constraints
- Need good heuristic for identifying potentially active constraints

- **Barrier/Interior Point (IP) Methods:**

- Add inequality constraints as “barriers” /penalties in the cost function. For example, $f'(\mathbf{x}) = f(\mathbf{x}) - \frac{1}{\rho} \log(-g(\mathbf{x}))$
- Very good for convex problems

What other methods are there?

- **Penalty Methods:**

- Add constraint violation into the cost function. For example,
$$f'(\mathbf{x}) = f(\mathbf{x}) + \frac{\rho}{2} \|\max(\mathbf{0}, g(\mathbf{x}))\|^2$$
- Similar to IP methods, but there is a key difference: IP methods blow up at the boundary, penalty methods start pushing only if there is a violation!
- Can easily get ill-conditioned
- Cannot get good accuracy

- **Sequential Quadratic Programming (SQP):**

- Sequential QP problems
- General idea:
 - Fit a quadratic approximation of the cost function
 - Linearize the constraints
 - Solve the produced QP with slack variables for feasibility
 - Iterate until convergence

Final Verdict on Optimization

- Newton's Method is very effective; Gauss-Newton the one used mostly in practice
- Line Search helps a lot
- Interior Point Methods are very effective in practice
- Transforming your problem into well known forms (e.g. QP) can greatly help
- A lot of good libraries

Thank you

- Any Questions?

- Office Hours:

- Tue & Thu (09:00-11:00)

- 24/7 by email (costashatz@upatras.gr, subject: *ECE_RSII_AM*)

- Material and Announcements



Laboratory of Automation & Robotics