

# ADMM

Daniil Merkulov

Optimization methods. MIPT

## Dual (sub)gradient method

Even if we can't derive dual (conjugate) in closed form, we can still use dual-based gradient or subgradient methods.

Consider the problem:

$$\begin{aligned} & \min_x f(x) \\ & \text{subject to } Ax = b \end{aligned}$$

### i Dual gradient ascent

$$\begin{aligned} x_k &\in \arg \min_x [f(x) + (u_{k-1})^T Ax] \\ u_k &= u_{k-1} + \alpha_k(Ax_k - b) \end{aligned}$$

- **Good:**  $x$  update decomposes when  $f$  does.

## Dual (sub)gradient method

Even if we can't derive dual (conjugate) in closed form, we can still use dual-based gradient or subgradient methods.

Consider the problem:

$$\begin{aligned} & \min_x f(x) \\ & \text{subject to } Ax = b \end{aligned}$$

### i Dual gradient ascent

$$\begin{aligned} x_k &\in \arg \min_x [f(x) + (u_{k-1})^T Ax] \\ u_k &= u_{k-1} + \alpha_k(Ax_k - b) \end{aligned}$$

- **Good:**  $x$  update decomposes when  $f$  does.
- **Bad:** require stringent assumptions (strong convexity of  $f$ ) to ensure convergence

## Augmented Lagrangian method aka method of multipliers

Augmented Lagrangian method transforms the primal problem to:

$$\begin{aligned} \min_x \quad & f(x) + \frac{\rho}{2} \|Ax - b\|^2 \\ \text{s.t. } & Ax = b \end{aligned}$$

## Augmented Lagrangian method aka method of multipliers

Augmented Lagrangian method transforms the primal problem to:

$$\begin{aligned} \min_x \quad & f(x) + \frac{\rho}{2} \|Ax - b\|^2 \\ \text{s.t. } & Ax = b \end{aligned}$$

where  $\rho > 0$  is a parameter. This formulation is clearly equivalent to the original problem. The problem is strongly convex if matrix  $A$  has full column rank.

## Augmented Lagrangian method aka method of multipliers

Augmented Lagrangian method transforms the primal problem to:

$$\begin{aligned} \min_x \quad & f(x) + \frac{\rho}{2} \|Ax - b\|^2 \\ \text{s.t. } & Ax = b \end{aligned}$$

where  $\rho > 0$  is a parameter. This formulation is clearly equivalent to the original problem. The problem is strongly convex if matrix  $A$  has full column rank.

**Dual gradient ascent:** The iterative updates are given by:

$$\begin{aligned} x_k &= \arg \min_x \left[ f(x) + (u_{k-1})^T Ax + \frac{\rho}{2} \|Ax - b\|^2 \right] \\ u_k &= u_{k-1} + \rho(Ax_k - b) \end{aligned}$$

↓ ogur  
zunefnep.

- **Good:** better convergence properties.

## Augmented Lagrangian method aka method of multipliers

Augmented Lagrangian method transforms the primal problem to:

$$\begin{aligned} \min_x \quad & f(x) + \frac{\rho}{2} \|Ax - b\|^2 \\ \text{s.t. } & Ax = b \end{aligned}$$

where  $\rho > 0$  is a parameter. This formulation is clearly equivalent to the original problem. The problem is strongly convex if matrix  $A$  has full column rank.

**Dual gradient ascent:** The iterative updates are given by:

$$\begin{aligned} x_k &= \arg \min_x \left[ f(x) + (u_{k-1})^T Ax + \frac{\rho}{2} \|Ax - b\|^2 \right] \\ u_k &= u_{k-1} + \rho(Ax_k - b) \end{aligned}$$

- **Good:** better convergence properties.
- **Bad:** lose decomposability

## Alternating Direction Method of Multipliers (ADMM)

Alternating direction method of multipliers or ADMM aims for the best of both worlds. Consider the following optimization problem:

Minimize the function:

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

$$\text{s.t. } Ax + Bz = c$$

ny emg  $x = z \Rightarrow$

$$\min f(x) + g(z)$$
$$I \cdot x - I \cdot z = 0$$

$$\min_{x} f(x) + g(x)$$

## Alternating Direction Method of Multipliers (ADMM)

**Alternating direction method of multipliers** or ADMM aims for the best of both worlds. Consider the following optimization problem:

Minimize the function:

$$\begin{aligned} & \min_{x,z} f(x) + g(z) \\ \text{s.t. } & Ax + Bz = c \end{aligned}$$

We augment the objective to include a penalty term for constraint violation:

$$\begin{aligned} & \min_{x,z} f(x) + g(z) + \frac{\rho}{2} \|Ax + Bz - c\|^2 \\ \text{s.t. } & Ax + Bz = c \end{aligned}$$

## Alternating Direction Method of Multipliers (ADMM)

**Alternating direction method of multipliers** or ADMM aims for the best of both worlds. Consider the following optimization problem:

Minimize the function:

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

$$\text{s.t. } Ax + Bz = c$$

We augment the objective to include a penalty term for constraint violation:

$$\min_{x,z} f(x) + g(z) + \frac{\rho}{2} \|Ax + Bz - c\|^2$$

$$\text{s.t. } Ax + Bz = c$$

where  $\rho > 0$  is a parameter. The augmented Lagrangian for this problem is defined as:

$$L_\rho(x, z, u) = f(x) + g(z) + u^T(Ax + Bz - c) + \frac{\rho}{2} \|Ax + Bz - c\|^2$$

## Alternating Direction Method of Multipliers (ADMM)

ADMM repeats the following steps, for  $k = 1, 2, 3, \dots$ :

1. Update  $x$ :

$$x_k = \arg \min_x L_\rho(x, z_{k-1}, u_{k-1})$$

## Alternating Direction Method of Multipliers (ADMM)

ADMM repeats the following steps, for  $k = 1, 2, 3, \dots$ :

1. Update  $x$ :

$$x_k = \arg \min_x L_\rho(x, z_{k-1}, u_{k-1})$$

2. Update  $z$ :

$$z_k = \arg \min_z L_\rho(x_k, z, u_{k-1})$$

# Alternating Direction Method of Multipliers (ADMM)

ADMM repeats the following steps, for  $k = 1, 2, 3, \dots$ :

1. Update  $x$ :

$$x_k = \arg \min_x L_\rho(x, z_{k-1}, u_{k-1})$$

2. Update  $z$ :

$$z_k = \arg \min_z L_\rho(x_k, z, u_{k-1})$$

3. Update  $u$ :

$$u_k = u_{k-1} + \rho(Ax_k + Bz_k - c)$$

# Alternating Direction Method of Multipliers (ADMM)

ADMM repeats the following steps, for  $k = 1, 2, 3, \dots$ :

1. Update  $x$ :

$$x_k = \arg \min_x L_\rho(x, z_{k-1}, u_{k-1})$$

2. Update  $z$ :

$$z_k = \arg \min_z L_\rho(x_k, z, u_{k-1})$$

3. Update  $u$ :

$$u_k = u_{k-1} + \rho(Ax_k + Bz_k - c)$$

# Alternating Direction Method of Multipliers (ADMM)

$$f(x,y) = xy$$

ADMM repeats the following steps, for  $k = 1, 2, 3, \dots$ :

1. Update  $x$ :

$$x_k = \arg \min_x L_\rho(x, z_{k-1}, u_{k-1})$$

$$f_x(y) = x \cdot y$$

2. Update  $z$ :

$$z_k = \arg \min_z L_\rho(x_k, z, u_{k-1})$$

$$f_y(x) = x \cdot y$$

3. Update  $u$ :

$$u_k = u_{k-1} + \rho(Ax_k + Bz_k - c)$$

Note: The usual method of multipliers would replace the first two steps by a joint minimization:

Matrix factorization

$$(x^{(k)}, z^{(k)}) = \arg \min_{x,z} L_\rho(x, z, u^{(k-1)})$$

ALS

Tensor Factorization

$$Y \ll \min(m, n)$$

$$X = U \cdot V$$

M<sub>m</sub>n

M<sub>r</sub>x<sub>r</sub>

$\rightarrow$

MM<sub>ΦΛ</sub>

$$\|X - UV\|_F^2 \rightarrow \min_{U,V}$$

# Convergence

Assume (very little!)

- $f, g$  convex, closed, proper

then ADMM converges:

**i** If the functions  $f$  and  $g$  are convex and computationally friendly for  $\arg \min$ , then ADMM has the following convergence bound for any  $x \in \mathbb{R}^{d_x}$ ,  $y \in \mathbb{R}^{d_y}$ ,  $\lambda \in \mathbb{R}^n$ :

$$L_0 \left( \frac{1}{k} \sum_{i=1}^k x_i, \frac{1}{k} \sum_{i=1}^K y_i, \lambda \right) - L_0(x, y, \frac{1}{k} \sum_{i=1}^k \lambda_k) \leq \frac{1}{2k} \|z_0 - z\|_P^2,$$

where  $L_0$  is the Lagrangian without augmentation,  $P$  and the initial value of  $z^0$  are defined as :

$$P = \begin{pmatrix} \rho A^T A & 0 & -A^T \\ 0 & 0 & 0 \\ -A & 0 & \frac{1}{\rho} I \end{pmatrix} \quad z^0 = \begin{pmatrix} x^0 \\ y^0 \\ \lambda^0 \end{pmatrix}.$$

# Convergence

Assume (very little!)

- $f, g$  convex, closed, proper
- $L_0$  has a saddle point

then ADMM converges:

**i** If the functions  $f$  and  $g$  are convex and computationally friendly for  $\arg \min$ , then ADMM has the following convergence bound for any  $x \in \mathbb{R}^{d_x}$ ,  $y \in \mathbb{R}^{d_y}$ ,  $\lambda \in \mathbb{R}^n$ :

$$L_0\left(\frac{1}{k} \sum_{i=1}^k x_i, \frac{1}{k} \sum_{i=1}^K y_i, \lambda\right) - L_0(x, y, \frac{1}{k} \sum_{i=1}^k \lambda_k) \leq \frac{1}{2k} \|z_0 - z\|_P^2,$$

where  $L_0$  is the Lagrangian without augmentation,  $P$  and the initial value of  $z^0$  are defined as :

$$P = \begin{pmatrix} \rho A^T A & 0 & -A^T \\ 0 & 0 & 0 \\ -A & 0 & \frac{1}{\rho} I \end{pmatrix} \quad z^0 = \begin{pmatrix} x^0 \\ y^0 \\ \lambda^0 \end{pmatrix}.$$

# Convergence

Assume (very little!)

- $f, g$  convex, closed, proper
- $L_0$  has a saddle point

then ADMM converges:

- iterates approach feasibility:  $Ax_k + Bz_k - c \rightarrow 0$

**i** If the functions  $f$  and  $g$  are convex and computationally friendly for  $\arg \min$ , then ADMM has the following convergence bound for any  $x \in \mathbb{R}^{d_x}$ ,  $y \in \mathbb{R}^{d_y}$ ,  $\lambda \in \mathbb{R}^n$ :

$$L_0\left(\frac{1}{k} \sum_{i=1}^k x_i, \frac{1}{k} \sum_{i=1}^K y_i, \lambda\right) - L_0(x, y, \frac{1}{k} \sum_{i=1}^k \lambda_i) \leq \frac{1}{2k} \|z_0 - z\|_P^2,$$

where  $L_0$  is the Lagrangian without augmentation,  $P$  and the initial value of  $z^0$  are defined as :

$$P = \begin{pmatrix} \rho A^T A & 0 & -A^T \\ 0 & 0 & 0 \\ -A & 0 & \frac{1}{\rho} I \end{pmatrix} \quad z^0 = \begin{pmatrix} x^0 \\ y^0 \\ \lambda^0 \end{pmatrix}.$$

# Convergence

Assume (very little!)

- $f, g$  convex, closed, proper
- $L_0$  has a saddle point

then ADMM converges:

- iterates approach feasibility:  $Ax_k + Bz_k - c \rightarrow 0$
- objective approaches optimal value:  $f(x_k) + g(z_k) \rightarrow p^*$

**i** If the functions  $f$  and  $g$  are convex and computationally friendly for  $\arg \min$ , then ADMM has the following convergence bound for any  $x \in \mathbb{R}^{d_x}$ ,  $y \in \mathbb{R}^{d_y}$ ,  $\lambda \in \mathbb{R}^n$ :

$$L_0 \left( \frac{1}{k} \sum_{i=1}^k x_i, \frac{1}{k} \sum_{i=1}^K y_i, \lambda \right) - L_0(x, y, \frac{1}{k} \sum_{i=1}^k \lambda_i) \leq \frac{1}{2k} \|z_0 - z\|_P^2,$$

where  $L_0$  is the Lagrangian without augmentation,  $P$  and the initial value of  $z^0$  are defined as :

$$P = \begin{pmatrix} \rho A^T A & 0 & -A^T \\ 0 & 0 & 0 \\ -A & 0 & \frac{1}{\rho} I \end{pmatrix} \quad z^0 = \begin{pmatrix} x^0 \\ y^0 \\ \lambda^0 \end{pmatrix}.$$

Example:

$$\min I_u(x) + I_v(z)$$

$$x - z = 0$$

$$0) L_g = I_u(x) + I_v(z) + \lambda^T(x - z) + \frac{\rho}{2} \|x - z\|_2^2$$

ADMM.

$$x_{k+1} = \underset{x}{\operatorname{argmin}} L_g(x, z, \lambda) =$$

$$= \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} I_u(x) + \lambda^T x + \frac{\rho}{2} \|x - z\|^2$$

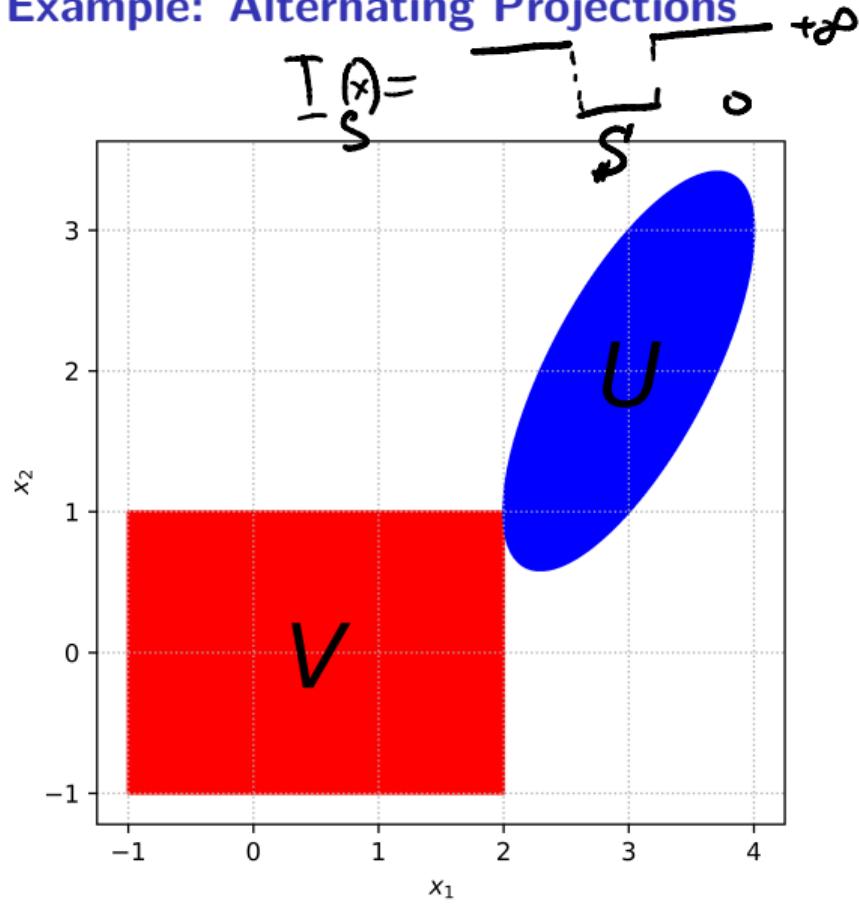
$$0 \in \partial I_u(x) + \lambda + \frac{\rho}{2}(x - z)$$

$$= \underset{x \in U}{\operatorname{argmin}} \frac{\rho}{2} \|x - z\|^2 + \lambda^T x$$

→ k шаг.

$$\underset{x \in U}{\operatorname{argmin}} \|x - \tilde{x}\|^2$$

## Example: Alternating Projections



Consider finding a point in the intersection of convex sets  $U, V \subseteq \mathbb{R}^n$ :

$$\min_x I_U(x) + I_V(x)$$

To transform this problem into ADMM form, we express it as:

$$\min_{x,z} I_U(x) + I_V(z) \quad \text{subject to } x - z = 0$$

Each ADMM cycle involves two projections:

$$x_k = \arg \min_x P_U(z_{k-1} - w_{k-1})$$

$$z_k = \arg \min_z P_V(x_k + w_{k-1})$$

$$w_k = w_{k-1} + x_k - z_k$$

$$\min f(x) + g(z)$$

$$\underline{Ax + Bz = C}$$

## Summary

- ADMM is one of the key and popular recent optimization methods.

## Summary

- ADMM is one of the key and popular recent optimization methods.
- It is implemented in many solvers and is often used as a default method.

## Summary

- ADMM is one of the key and popular recent optimization methods.
- It is implemented in many solvers and is often used as a default method.
- The non-standard formulation of the problem itself, for which ADMM is invented, turns out to include many important special cases. “Unusual” variable  $y$  often plays the role of an auxiliary variable.

## Summary

- ADMM is one of the key and popular recent optimization methods.
- It is implemented in many solvers and is often used as a default method.
- The non-standard formulation of the problem itself, for which ADMM is invented, turns out to include many important special cases. “Unusual” variable  $y$  often plays the role of an auxiliary variable.
- Here the penalty is an additional modification to stabilize and accelerate convergence. It is not necessary to make  $\rho$  very large.

## Sources

- Ryan Tibshirani. Convex Optimization 10-725