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:

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

Dual gradient ascent

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

- **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:

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

Dual gradient ascent

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

- **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:

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

Augmented Lagrangian method aka method of multipliers

Augmented Lagrangian method transforms the primal problem to:

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

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:

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

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.

Augmented Lagrangian method aka method of multipliers

Augmented Lagrangian method transforms the primal problem to:

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

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:

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

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

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:

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

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

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 $k = 1, 2, 3, \ldots$:

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)

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

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:

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

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

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

Convergence

Assume (very little!)

- f, g convex, closed, proper

then ADMM converges:

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

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

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:

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: $Ax_k + Bz_k - c \to 0$

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

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

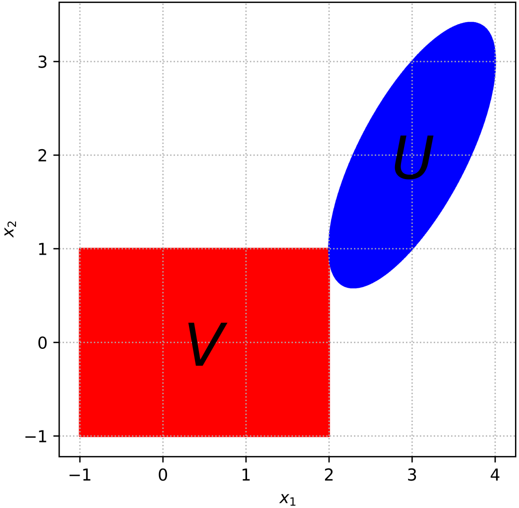
$$L_0\left(\frac{1}{k}\sum_{i=1}^{k} x_i, \frac{1}{k}\sum_{i=1}^{K} y_i, \lambda\right) - L_0(x, y, \frac{1}{k}\sum_{i=1}^{k}\lambda_k) \leq \frac{1}{2k}\left\|z_0 - z\right\|_P^2,$$

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:

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:

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

Each ADMM cycle involves two projections:

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

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

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

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 ρ very large.

Sources

- Alternating Direction Method of Multipliers by S.Boyd

Sources

- Alternating Direction Method of Multipliers by S.Boyd
- Ryan Tibshirani. ConvAlternating Direction Method of Multipliers by S.Boydex Optimization 10-725