Anderson Accelerated Douglas-Rachford Splitting

Anqi Fu Junzi Zhang Stephen Boyd

EE & ICME Departments

Stanford University

February 26, 2020

Outline

Prox-Affine Form

Prox-affine convex optimization problem:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^N f_i(x_i) \\ \text{subject to} & \sum_{i=1}^N A_i x_i = b \end{array}$$

with variables $x_i \in \mathbf{R}^{n_i}$ for $i = 1, \dots, N$

- $A_i \in \mathbf{R}^{m \times n_i}$ and $b \in \mathbf{R}^m$ given data
- $f_i : \mathbf{R}^{n_i} \to \mathbf{R} \cup \{+\infty\}$ are closed, convex and proper
- Each f_i can only be accessed via its proximal operator

$$\mathbf{prox}_{tf_i}(v_i) = \text{argmin}_{x_i} \left\{ f_i(x_i) + \frac{1}{2t} \|x_i - v_i\|_2^2 \right\},$$

where $t > 0$ is a parameter

Why This Formulation?

- Encompasses many classes of convex problems (conic programs, consensus optimization)
- Block separable form ideal for distributed optimization
- Proximal operator can be provided as a “black box”, enabling privacy-preserving implementation

Previous Work

- Alternating direction method of multipliers (ADMM)
- Douglas-Rachford splitting (DRS)
- Augmented Lagrangian method (ALM)

Previous Work

- Alternating direction method of multipliers (ADMM)
- Douglas-Rachford splitting (DRS)
- Augmented Lagrangian method (ALM)

These are typically slow to converge, prompting research into acceleration techniques:

- Adaptive penalty parameters
- Momentum methods
- Quasi-Newton method with line search

Our Method

- **A2DR**: Anderson acceleration (AA) applied to DRS
- DRS is a non-expansive fixed-point (NEFP) method that fits prox-affine framework
- AA is fast, efficient, and can be applied to NEFP iterations – but unstable without modification
- We introduce a type-II AA variant that converges globally in non-smooth, potentially pathological settings

Main Advantages

- A2DR produces primal and dual solutions, or a certificate of infeasibility/unboundedness
- Consistently converges faster with no parameter tuning
- Memory efficient $\Rightarrow$ little extra cost per iteration
- Scales to large problems and is easily parallelized
- Python implementation:

  `https://github.com/cvxgrp/a2dr`

Outline

DRS Algorithm

- Rewrite problem as

$$\text{minimize} \quad \sum_{i=1}^{N} f_i(x_i) + \mathcal{I}_{Ax=b}(x),$$

where $\mathcal{I}_S$ is the indicator of set S

- DRS iterates for $k = 1, 2, \ldots,$

$$\begin{aligned} x_i^{k+1/2} &= \mathbf{prox}_{tf_i}(v^k), \quad i = 1, \ldots, N \\ v^{k+1/2} &= 2x^{k+1/2} - v^k \\ x^{k+1} &= \Pi_{Av=b}(v^{k+1/2}) \\ v^{k+1} &= v^k + x^{k+1} - x^{k+1/2} \end{aligned}$$

$\Pi_S(v)$ is Euclidean projection of v onto S

Convergence of DRS

- DRS iterations can be conceived as a fixed-point mapping

$$v^{k+1} = F(v^k),$$

  where F is firmly non-expansive
- v^k converges to a fixed point of F (if it exists)
- x^k and $x^{k+1/2}$ converge to a solution of our problem

Convergence of DRS

- DRS iterations can be conceived as a fixed-point mapping

$$v^{k+1} = F(v^k),$$

  where F is firmly non-expansive
- v^k converges to a fixed point of F (if it exists)
- x^k and $x^{k+1/2}$ converge to a solution of our problem

In practice, this convergence is often slow...

Outline

Type-II AA

- Quasi-Newton method for accelerating fixed point iterations
- **Extrapolates** next iterate using $M+1$ most recent iterates

$$v^{k+1} = \sum_{j=0}^{M} \alpha_j^k F(v^{k-M+j})$$

- Let $G(v) = v - F(v)$, then $\alpha^k \in \mathbf{R}^{M+1}$ is solution to

$$\begin{array}{ll} \text{minimize} & \|\sum_{j=0}^{M} \alpha_j^k G(v^{k-M+j})\|_2^2 \\ \text{subject to} & \sum_{j=0}^{M} \alpha_j^k = 1 \end{array}$$

- Typically only need $M \approx 10$ for good performance

Adaptive Regularization

- Type-II AA is unstable so we add a regularization term
- Change variables to $\gamma^k \in \mathbf{R}^M$

$$\alpha_0^k = \gamma_0^k, \quad \alpha_i^k = \gamma_i^k - \gamma_{i-1}^k \; \forall i = 1, \ldots, M-1, \quad \alpha_M^k = 1 - \gamma_{M-1}^k$$

- Stabilized AA problem is

$$\text{minimize} \quad \|g^k - Y_k \gamma^k\|_2^2 + \eta \left(\|S_k\|_F^2 + \|Y_k\|_F^2\right) \|\gamma^k\|_2^2,$$

where $\eta \geq 0$ is a parameter and

$$g^k = G(v^k), \quad y^k = g^{k+1} - g^k, \quad Y_k = [y^{k-M} \; \ldots \; y^{k-1}]$$
$$s^k = v^{k+1} - v^k, \quad S_k = [s^{k-M} \; \ldots \; s^{k-1}]$$

A2DR

- Let $\alpha = H(v, g)$ be the weights produced by stabilized AA
- A2DR iterates for $k = 1, 2, \ldots,$

$$
\begin{aligned}
v_{\text{DRS}}^{k+1} &= F(v^k), \quad g^k = v^k - v_{\text{DRS}}^{k+1} \\
\alpha^k &= H(v^k, g^k) \\
v_{\text{AA}}^{k+1} &= \sum_{j=0}^{M} \alpha_j^k v_{\text{DRS}}^{k-M+j+1} \\
v^{k+1} &= \begin{cases} v_{\text{AA}}^{k+1} & \text{safeguard passes} \\ v_{\text{DRS}}^{k+1} & \text{safeguard fails} \end{cases}
\end{aligned}
$$

Stopping Criterion of A2DR

- Stop and output $x^{k+1/2}$ when $\|r^k\|_2 \leq \epsilon_{\text{tol}}$

$$r^k_{\text{prim}} = Ax^{k+1/2} - b$$
$$r^k_{\text{dual}} = \frac{1}{t}(v^k - x^{k+1/2}) + A^T\lambda^k$$

- Dual variable is solution to least-squares problem

$$\lambda^k = \operatorname{argmin} \|r^k_{\text{dual}}\|_2$$

Convergence of A2DR

Theorem (Solvable Case)

If the problem is feasible and bounded,

$$\liminf_{k\to\infty} \|r^k\|_2 = 0$$

and the AA candidates are adopted infinitely often. Furthermore, if F has a fixed point $v^\star$,

$$\lim_{k\to\infty} v^k = v^\star \text{ and } \lim_{k\to\infty} x^{k+1/2} = x^\star,$$

where $x^\star$ is a solution to the problem.

Convergence of A2DR

Theorem (Pathological Case)

If the problem is pathological,

$$\lim_{k \to \infty} \left(v^k - v^{k+1} \right) = \delta v \neq 0.$$

Furthermore, if $\lim_{k \to \infty} A x^{k+1/2} = b$*, the problem is unbounded. Otherwise, it is infeasible.*

Preconditioning

- Convergence greatly improved by rescaling problem
- Replace original A, b, f_i with

$$\hat{A} = DAE, \quad \hat{b} = Db, \quad \hat{f}_i(\hat{x}_i) = f_i(e_i \hat{x}_i)$$

- D and E are diagonal positive, $e_i > 0$ corresponds to ith block diagonal entry of E
- D and E chosen by equilibrating A (see paper for details)
- Proximal operator of $\hat{f}_i$ can be evaluated using proximal operator of f_i

$$\mathbf{prox}_{t\hat{f}_i}(\hat{v}_i) = \frac{1}{e_i}\mathbf{prox}_{(e_i^2 t)f_i}(e_i \hat{v}_i)$$

Outline

Nonnegative Least Squares (NNLS)

$$\begin{array}{ll} \text{minimize} & \|Fz - g\|_2^2 \\ \text{subject to} & z \geq 0 \end{array}$$

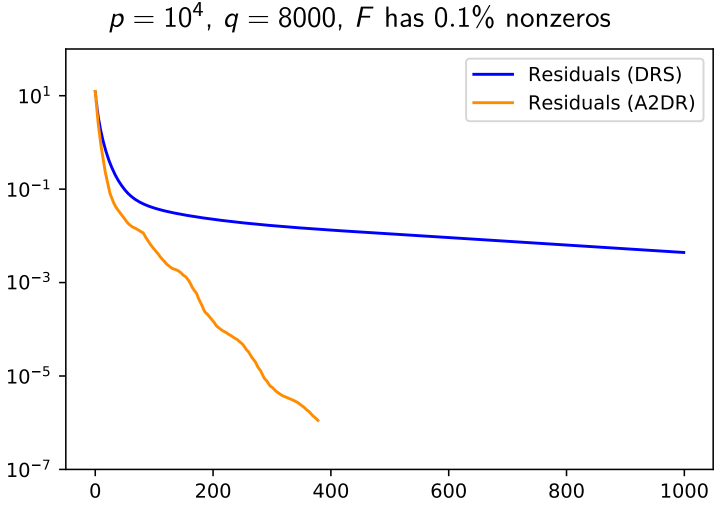
with respect to $z \in \mathbf{R}^q$

- Problem data: $F \in \mathbf{R}^{p \times q}$ and $g \in \mathbf{R}^p$
- Can be written in standard form with

$$f_1(x_1) = \|Fx_1 - g\|_2^2, \quad f_2(x_2) = \mathcal{I}_{\mathbf{R}_+^n}(x_2)$$
$$A_1 = I, \quad A_2 = -I, \quad b = 0$$

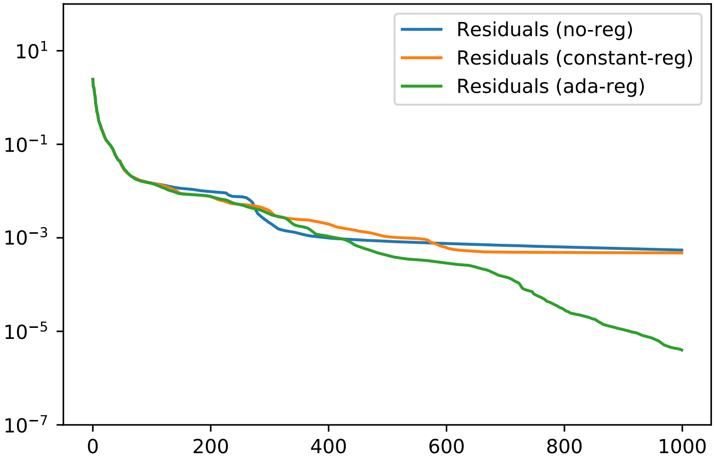
- We evaluate proximal operator of f_1 using LSQR

NNLS: Convergence of $\|r^k\|_2$



NNLS: Convergence of $\|r^k\|_2$

$p = 300$, $q = 500$, F has 0.1% nonzeros



Sparse Inverse Covariance Estimation

- Samples $z_1, \ldots, z_p$ IID from $\mathcal{N}(0, \Sigma)$
- Know covariance $\Sigma \in \mathbf{S}_+^q$ has **sparse** inverse $S = \Sigma^{-1}$
- One way to estimate S is by solving the penalized log-likelihood problem

$$\text{minimize} \quad -\log\det(S) + \mathrm{tr}(SQ) + \alpha\|S\|_1,$$

  where Q is the sample covariance, $\alpha \geq 0$ is a parameter
- Note $\log\det(S) = -\infty$ when $S \not\succ 0$

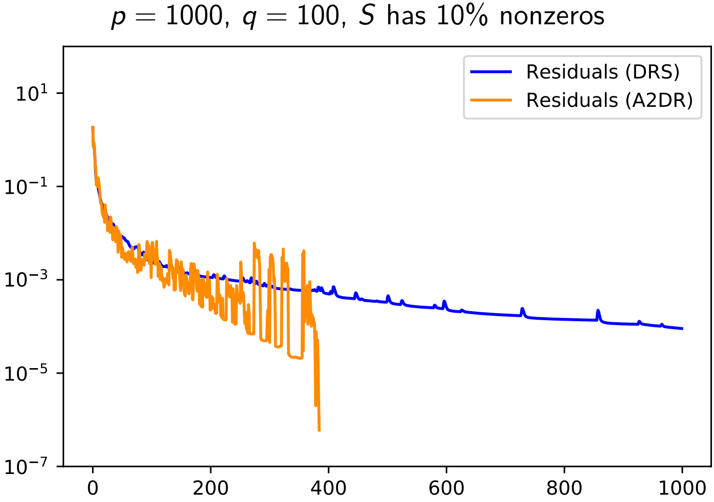
Sparse Inverse Covariance Estimation

- Problem can be written in standard form with

$$f_1(S_1) = -\log\det(S_1) + \mathrm{tr}(S_1 Q), \quad f_2(S_2) = \alpha\|S_2\|_1$$
$$A_1 = I, \quad A_2 = -I, \quad b = 0$$

- Both proximal operators have closed-form solutions (Parikh & Boyd 2014)

Covariance Estimation: Convergence of $\|r^k\|_2$



Multi-Task Logistic Regression

$$\text{minimize} \quad \phi(W\theta, Y) + \alpha \textstyle\sum_{l=1}^{L} \|\theta_l\|_2 + \beta\|\theta\|_*$$

with respect to $\theta = [\theta_1 \cdots \theta_L] \in \mathbf{R}^{s \times L}$

- Problem data: $W \in \mathbf{R}^{p \times s}$ and $Y = [y_1 \cdots y_L] \in \mathbf{R}^{p \times L}$
- Regularization parameters: $\alpha \geq 0, \beta \geq 0$
- Logistic loss function

$$\phi(Z, Y) = \sum_{l=1}^{L} \sum_{i=1}^{p} \log\left(1 + \exp(-Y_{il} Z_{il})\right)$$

Multi-Task Logistic Regression

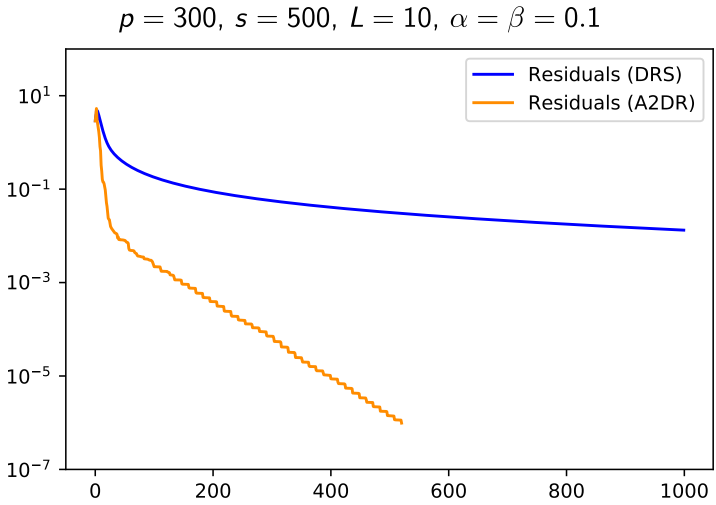
- Rewrite problem in standard form with

$$f_1(Z) = \phi(Z, Y), \quad f_2(\theta) = \alpha \sum_{l=1}^{L} \|\theta_l\|_2, \quad f_3(\tilde{\theta}) = \beta \|\tilde{\theta}\|_*,$$

$$A = \begin{bmatrix} I & -W & 0 \\ 0 & I & -I \end{bmatrix}, \quad x = \begin{bmatrix} Z \\ \theta \\ \tilde{\theta} \end{bmatrix}, \quad b = 0$$

- We evaluate proximal operator of f_1 using Newton-CG method, rest have closed-form solutions

Multi-Task Logistic: Convergence of $\|r^k\|_2$



Outline

Conclusion

- A2DR is a fast, robust algorithm for solving linearly constrained convex optimization problems
- Can be easily scaled up and parallelized
- Open-source Python solver:

  `https://github.com/cvxgrp/a2dr`

Future Work

- More work on feasibility detection
- Expand library of proximal operators
- User-friendly interface with CVXPY
- GPU parallelization and cloud computing