Trajectory of Alternating Direction Method of Multipliers and Adaptive Acceleration

Clarice Poon (University of Bath) Jingwei Liang (University of Cambridge)

Alternating direction method of multipliers (ADMM)

Consider the minimization problem

$$\min_{x\in\mathbb{R}^n, y\in\mathbb{R}^m} R(x)+J(y) \quad \text{such that} \quad Ax+By=b, \qquad (\mathcal{P}_{\mathrm{ADMM}})$$

where the following basic assumptions are imposed

- $R\in\Gamma_0(\mathbb{R}^n)$, $J\in\Gamma_0(\mathbb{R}^m)$ are proper closed and convex functions.
- $A:\mathbb{R}^n\to\mathbb{R}^p$ and $B:\mathbb{R}^m\to\mathbb{R}^p$ are injective linear operators.
- $\mathrm{ri}(\mathrm{dom}(R)\cap\mathrm{dom}(J))\neq\emptyset$, and the set of minimizers is non-empty.

Augmented Lagrangian associated to $(\mathcal{P}_{\mathrm{ADMM}})$

$$\mathcal{L}(x,y;\psi)\overset{\text{def}}{=}R(x)+J(y)+\langle\psi, Ax+By-b\rangle+\tfrac{\gamma}{2}\|Ax+By-b\|^2.$$

Alternating direction method of multipliers

$$\begin{aligned} x_k &= \mathrm{argmin}_{x\in\mathbb{R}^n} R(x)+\tfrac{\gamma}{2}\|Ax+By_{k-1}-b+\tfrac{1}{\gamma}\psi_{k-1}\|^2,\\ y_k &= \mathrm{argmin}_{y\in\mathbb{R}^m} J(y)+\tfrac{\gamma}{2}\|Ax_k+By-b+\tfrac{1}{\gamma}\psi_{k-1}\|^2,\\ \psi_k &= \psi_{k-1}+\gamma(Ax_k+By_k-b).\end{aligned}$$

Define $z_k\overset{\text{def}}{=}\psi_{k-1}+\gamma Ax_k$, we can rewrite ADMM as

$$\begin{aligned} x_k &= \mathrm{argmin}_{x\in\mathbb{R}^n} R(x)+\tfrac{\gamma}{2}\|Ax-\tfrac{1}{\gamma}(z_{k-1}-2\psi_{k-1})\|^2,\\ z_k &= \psi_{k-1}+\gamma Ax_k,\\ y_k &= \mathrm{argmin}_{y\in\mathbb{R}^m} J(y)+\tfrac{\gamma}{2}\|By+\tfrac{1}{\gamma}(z_k-\gamma b)\|^2,\\ \psi_k &= z_k+\gamma(By_k-b).\end{aligned}$$

Fixed-point characterization: there exists some $\mathcal{F}$ such that

$$z_{k+1}=\mathcal{F}(z_k).$$

Trajectory of ADMM and failure of inertial

Linearization For k large enough

$$z_{k+1}-z_k=M(z_k-z_{k-1})+o(\|z_k-z_{k-1}\|).$$

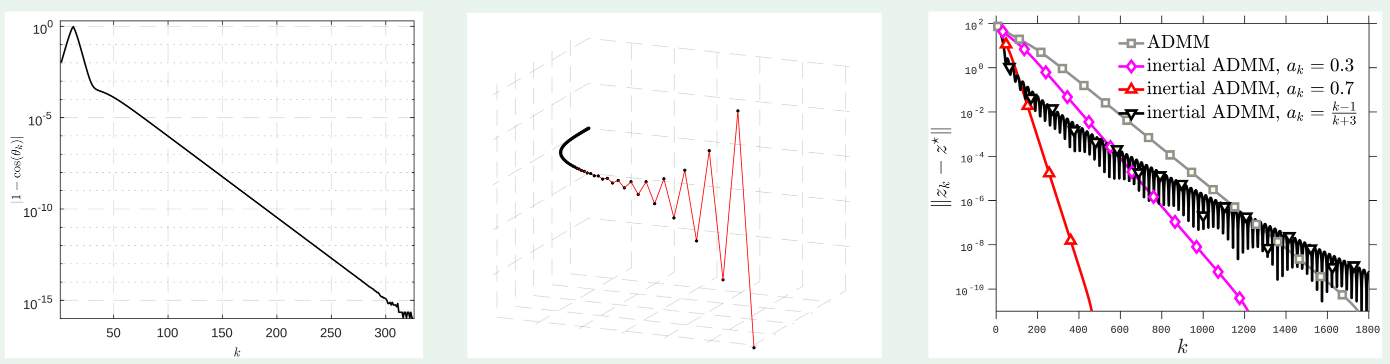
Define $v_k=z_k-z_{k-1}$ and $\theta_k=\angle(v_k,v_{k-1})$.

LASSO problem

$$\min_{x,y\in\mathbb{R}^n} \mu\|x\|_1+\tfrac{1}{2}\|Ky-f\|^2 \quad \text{such that} \quad x-y=0.$$

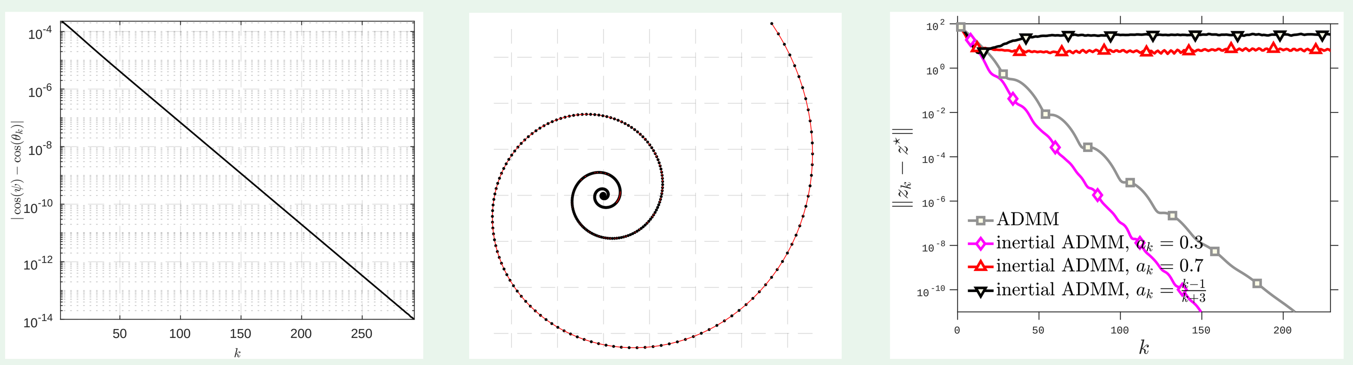
Trajectory of z_k

Straight-line trajectory: $\theta_k\to 0$



$\theta_k\to 0$ | Trajectory of $\{z_k\}_{k\in\mathbb{N}}$ | Success of inertial

Spiral trajectory: $\theta_k\to\theta^\star\in]0,\pi/2[$



$\theta_k\to\theta^\star$ | Trajectory of $\{z_k\}_{k\in\mathbb{N}}$ | **Failure** of inertial

Trajectory based Adaptive Acceleration

The regularity of trajectory allows to use the current points to predict the future points. That is

$$\bar{z}_{k,s}=\mathcal{E}_{s,q}(z_k,z_{k-1},\cdots,z_{k-q}).$$

Idea: given $\{z_{k-j}\}_{j=0}^{q+1}$ and $v_{k-j}\overset{\text{def}}{=}z_{k-j}-z_{k-j-1}$, predict the future iterates by considering how the past directions $v_{k-1},\dots,v_{k-q}$ approximate the latest direction v_k:

- Let $V_{k-1}\overset{\text{def}}{=}[v_{k-1},\cdots,v_{k-q}]\in\mathbb{R}^{n\times q}$, and
$$c_k\overset{\text{def}}{=}\mathrm{argmin}_{c\in\mathbb{R}^q}\|V_{k-1}c-v_k\|^2=\|\textstyle\sum_{j=1}^q c_jv_{k-j}-v_k\|^2.$$
- The idea is then that $v_{k+1}\approx V_kc_k$ and so, $\bar{z}_{k,1}\overset{\text{def}}{=}z_k+V_kc\approx z_{k+1}$. Iterating this s times, we obtain $\bar{z}_{k,s}\approx z_{k+s}$.

Given $c\in\mathbb{R}^q$, define the mapping H by

$$H(c)=\left[\begin{array}{c|c} c_{1:q-1} & \mathrm{Id}_{q-1}\\ c_q & 0_{1,q-1}\end{array}\right]\in\mathbb{R}^{q\times q}.$$

Let $C_k=H(c_k)$, note that $V_k=V_{k-1}C_k$. Let $(C)_{(:,1)}$ be the first column of C, then

$$\bar{z}_{k,s}=z_k+V_k(\textstyle\sum_{i=1}^s C_k^i)_{(:,1)},$$

which is the desired trajectory following extrapolation scheme.

Algorithm: A^{3}DMM - adaptive acceleration for ADMM

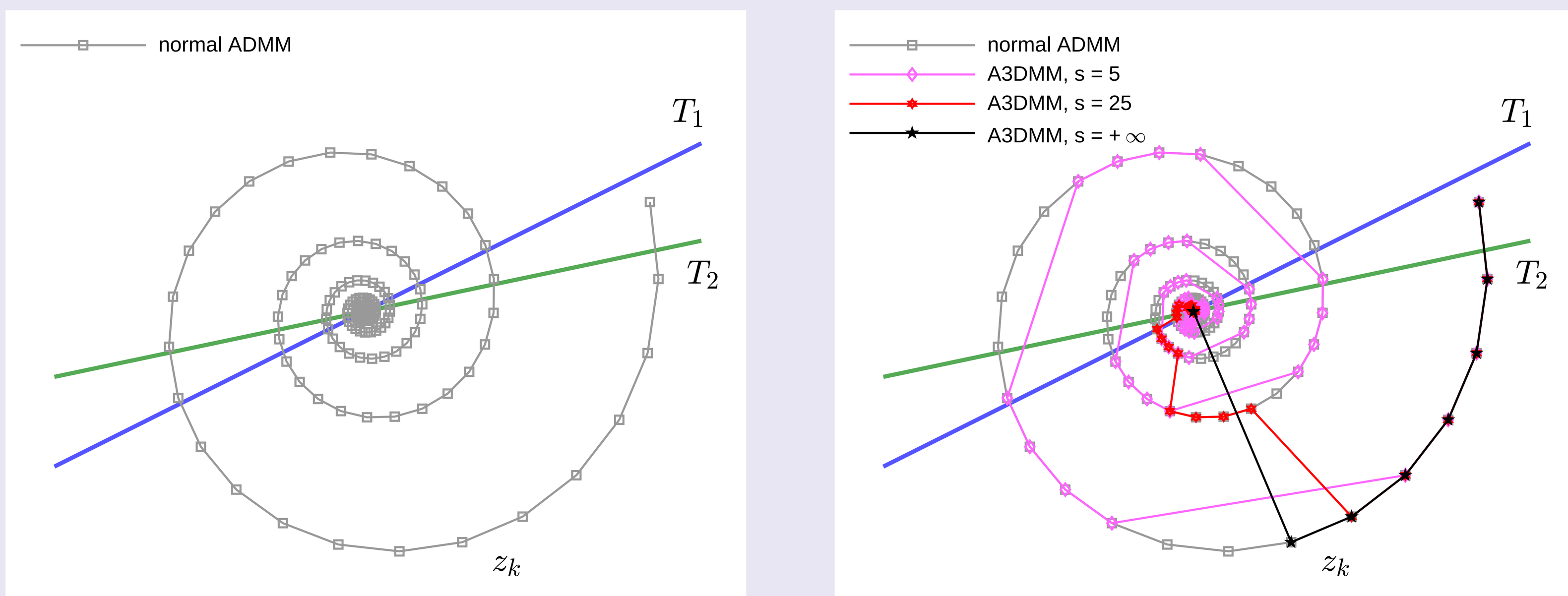
Initial: Let $s\geq 1, q\geq 1$ be integers, $V_0=0\in\mathbb{R}^{p\times(q+1)}$;
Repeat:

- For $k\geq 1$:
$$\begin{aligned} y_k &= \mathrm{argmin}_{y\in\mathbb{R}^m} J(y)+\tfrac{\gamma}{2}\|By+\tfrac{1}{\gamma}(\bar{z}_{k-1}-\gamma b)\|^2,\\ \psi_k &= \bar{z}_{k-1}+\gamma(By_k-b),\\ x_k &= \mathrm{argmin}_{x\in\mathbb{R}^n} R(x)+\tfrac{\gamma}{2}\|Ax-\tfrac{1}{\gamma}(\bar{z}_{k-1}-2\psi_k)\|^2,\\ z_k &= \psi_k+\gamma Ax_k,\\ v_k &= z_k-z_{k-1} \quad\text{and}\quad V_k=[v_k,V_{k-1}(:,1:q-1)].\end{aligned}$$
- If $\mathrm{mod}(k,q+2)=0$: compute $C_k=H(c_k)$, if $\rho(C_k)<1$: $\bar{z}_k=z_k+V_k(\sum_{i=1}^s C_k^i)_{(:,1)}$.
- If $\mathrm{mod}(k,q+2)\neq 0$: $\bar{z}_k=z_k$.

Until: $\|v_k\|\leq\mathrm{tol}$.

- When $s=+\infty$, $\bar{z}_{k,\infty}=\frac{1}{1-\sum_{i=1}^s c_{k,i}}(z_k-\sum_{j=1}^{q-1}c_{k,j}z_{k-j})$ which is equivalent to *minimal polynomial extrapolation* (MPE) but using **different past points**.
- $\|\bar{z}_{k,s}-z^\star\|\leq\|z_{k+s}-z^\star\|+B_s\epsilon_k$ where $\epsilon_k\overset{\text{def}}{=}\|V_{k-1}c-v_k\|=\mathcal{O}(|\lambda_{q+1}|^k)$.

Example



Feasibility problem and ADMM | A^{3}DMM

- In $\mathbb{R}^2$, the trajectory of $\{z_k\}_{k\in\mathbb{N}}$ is a logarithmic spiral.
- A^{3}DMM with $q=2$ and $s=5,25,+\infty$.

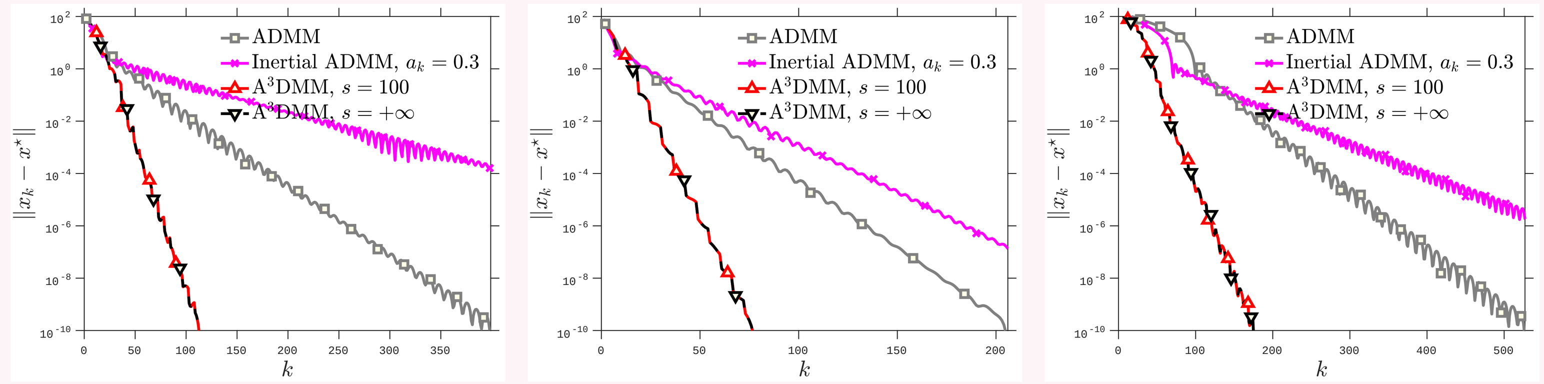
- A. S. Lewis. Active sets, non-smoothness, and sensitivity. SIAM Journal on Optimization, 13(3):702–725, 2003.
- A. Sidi. Vector extrapolation methods with applications, volume 17. SIAM, 2017.
- S. Cabay and L. Jackson. A polynomial extrapolation method for finding limits and anti-limits of vector sequences. SIAM Journal on Numerical Analysis, 13(5):734–752, 1976.

Numerical Experiments

We fix $q=4$ and two choices of s are considered: $s=100$ and $s=+\infty$.

Affine constrained minimization

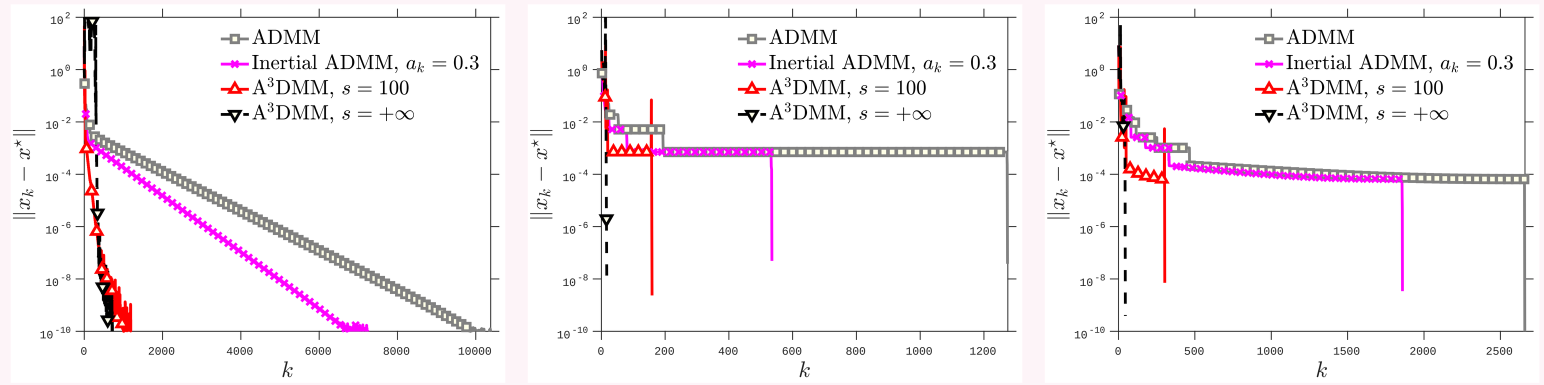
$$\min_{x,y\in\mathbb{R}^n} R(x)+\iota_{\{y:Ky=f\}}(y) \quad\text{such that}\quad x-y=0.$$



(a) ℓ_1-norm: $\|x_k-x^\star\|$ (b) $\ell_{1,2}$-norm: $\|x_k-x^\star\|$ (c) Nuclear norm: $\|x_k-x^\star\|$

LASSO

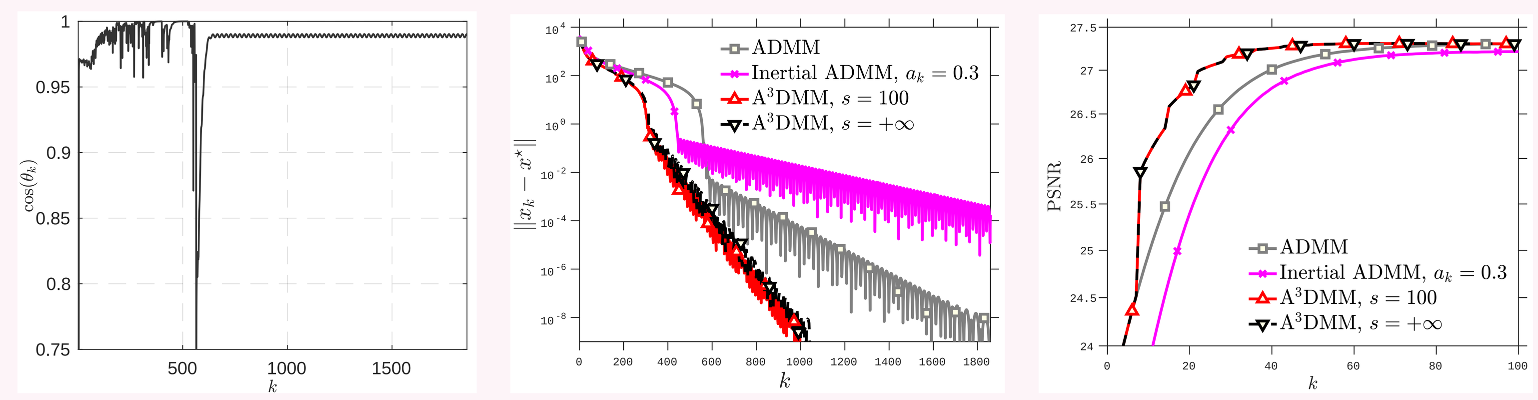
$$\min_{x,y\in\mathbb{R}^n} R(x)+\tfrac{1}{2}\|Ky-f\|^2 \quad\text{such that}\quad x-y=0.$$



(d) covtype: $\|x_k-x^\star\|$ (e) ijcnn1: $\|x_k-x^\star\|$ (f) phishing: $\|x_k-x^\star\|$

Total variation image inpainting

$$\min_{x\in\mathbb{R}^{n\times n}, y\in\mathbb{R}^{2n\times n}} \|y\|_1+\iota_{\{x:\mathcal{P}_\Omega x=f\}}(x) \quad\text{such that}\quad \nabla x-y=0.$$

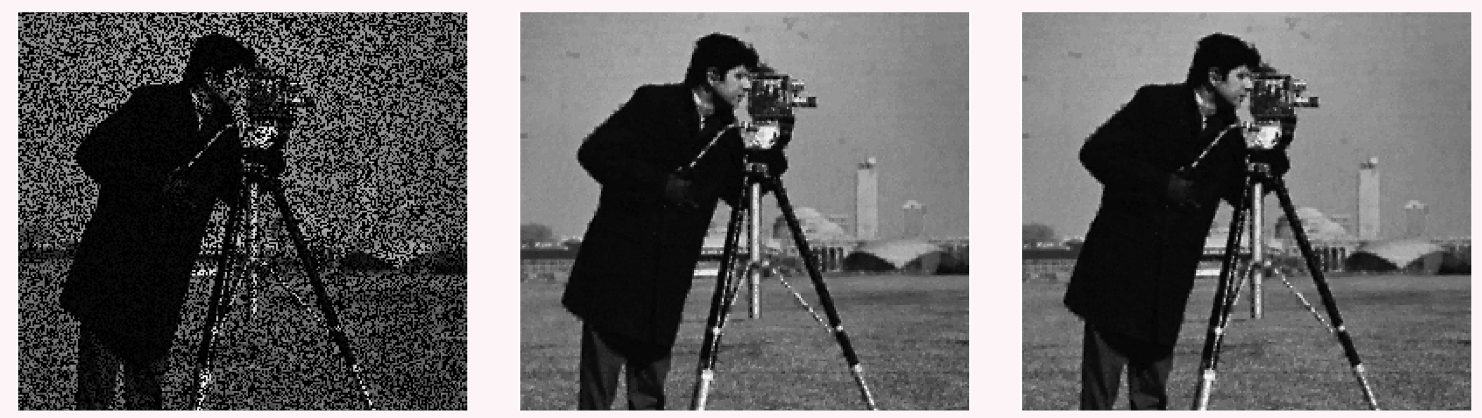


(g) Angle $\{\theta_k\}_{k\in\mathbb{N}}$ of ADMM (h) Comparison of $\|x_k-x^\star\|$ (i) PSNR value

NB: oscillatory $\cos(\theta_k)$ due to subproblem x_k is solved approximately.

Image quality comparison at iteration step $k=30$:

(j) Original image (k) ADMM, PSNR = 26.5448 (l) Inertial ADMM, PSNR = 26.1096



(m) Observed image (n) A^{3}DMM $s=100$, PSNR = 27.0402 (o) A^{3}DMM $s=+\infty$, PSNR = 27.0402

cmhsp20@bath.ac.uk; jl993@cam.ac.uk