Accelerated Primal Dual Fixed Point Method

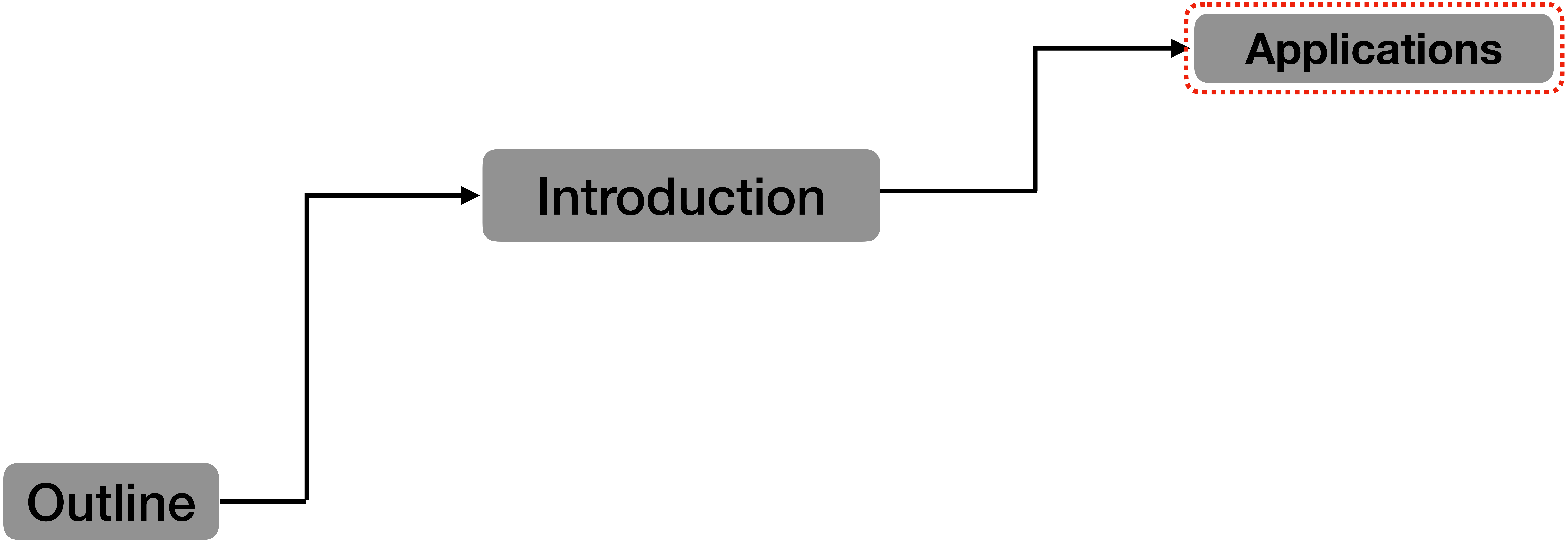
Ya-Nan Zhu （朱亚南）

Union of Mathematical Imaging（UMI）

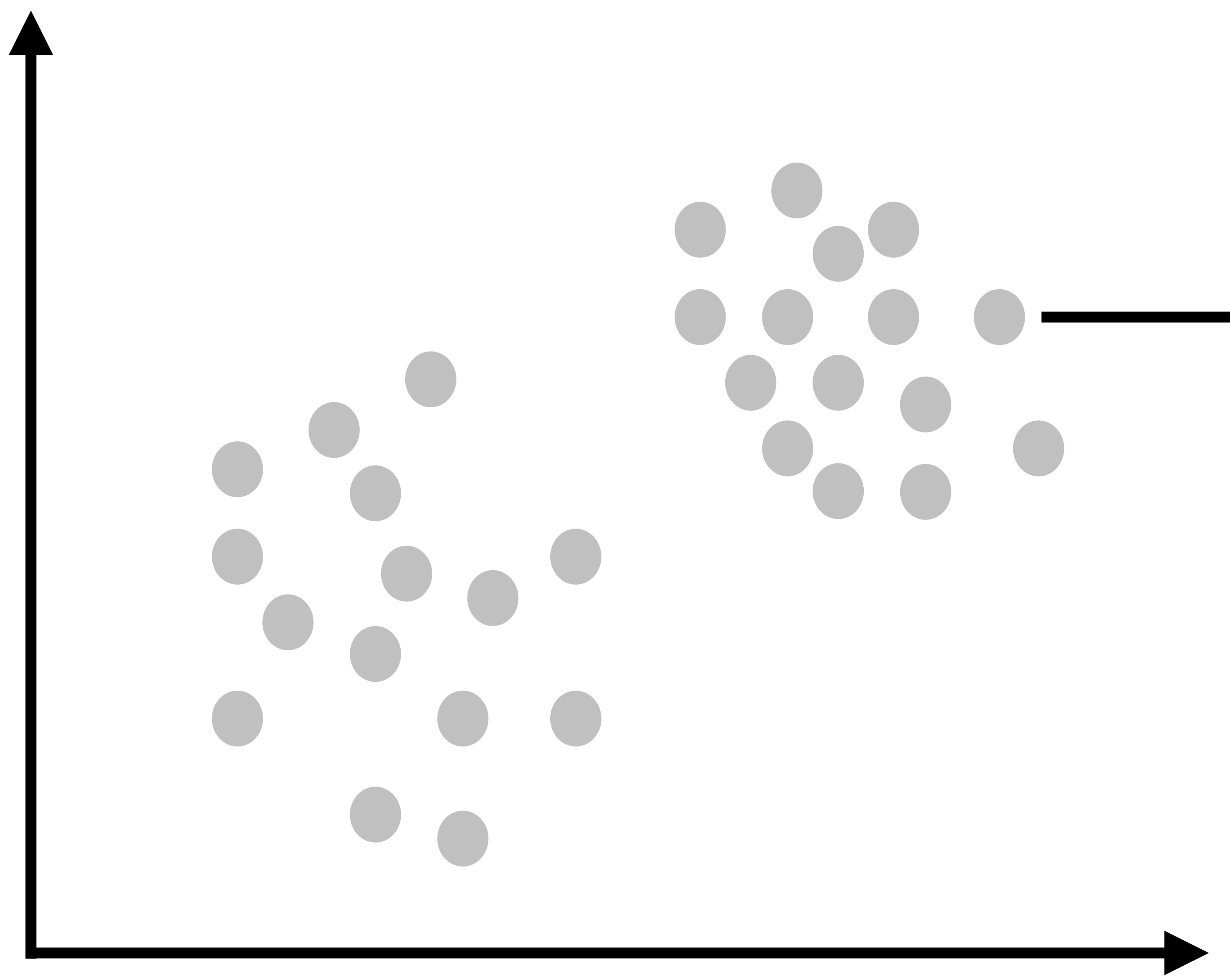
September. 19-21, 2025
Enshi, Hubei, China

哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

Outline
Introduction
Applications
Mathematical Model
Algorithms Review
Motivation
Algorithm
APDFP
Convergence Results
Numerical Experiments
Conclusions


Applications
Introduction
Outline

Data Sciences



(a_i, b_i)

sample size

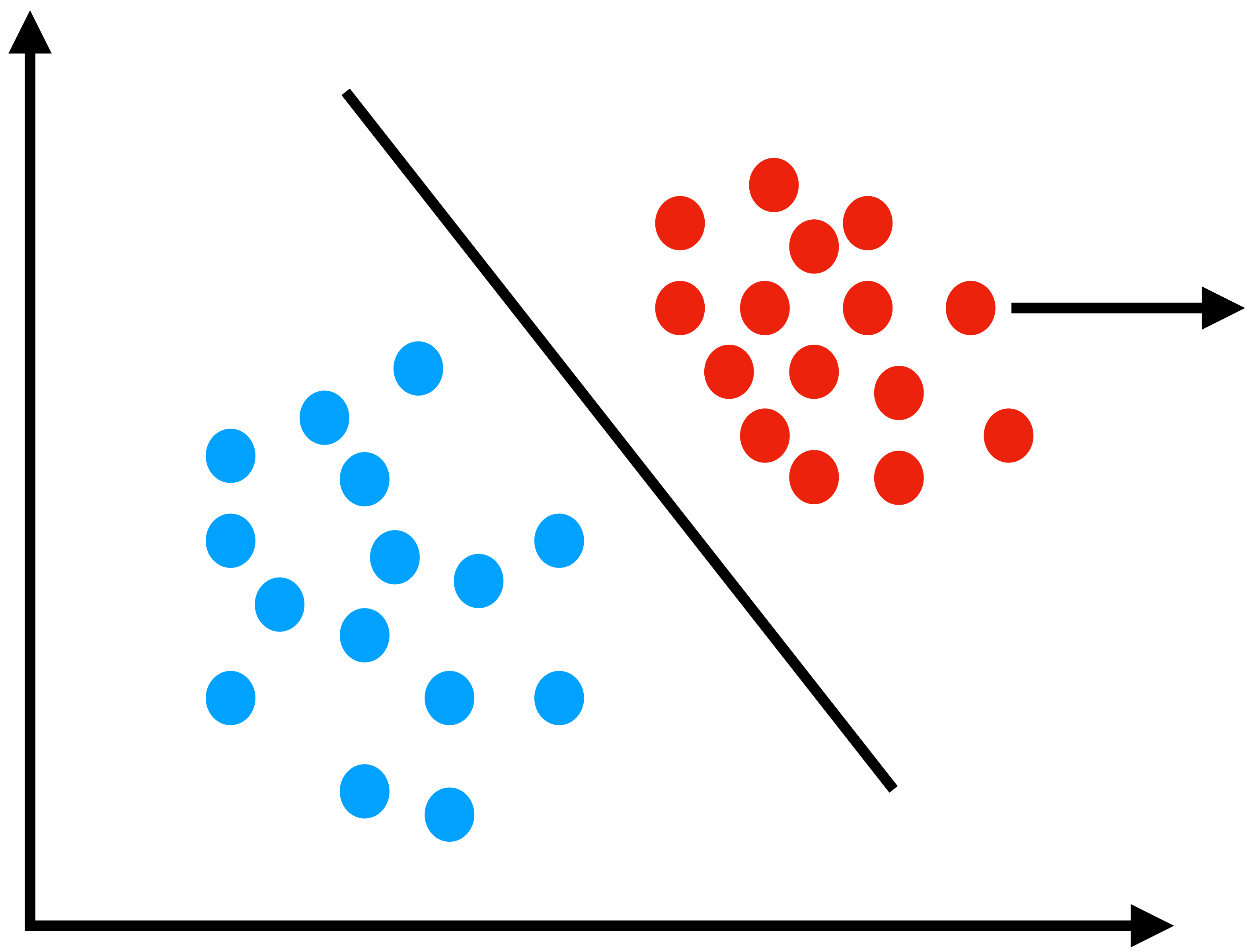
$$x^* = \arg\min_{x\in\mathbb{R}^d} \frac{1}{n}\sum_{i=1}^{n} l_i(x, a_i, b_i)$$

Model weight

Loss function

Linear Regression	Logistic Regression	Support Vector Machine
Squared Loss	Logistic Loss	Hinge Loss
$l_i(x, a_i, b_i) = (a_i^T x - b_i)^2$	$l_i(x, a_i, b_i) = \log(1 + \exp(-b_i a_i^T x))$	$l_i(x, a_i, b_i) = \max\{0, 1 - b_i a_i^T x\}$

Data Sciences



(a_i, b_i)

sample size

$$x^* = \arg\min_{x \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \boxed{l_i(x, \boxed{a_i}, \boxed{b_i})}$$

Model weight

Loss function

Linear Regression

Squared Loss

$$l_i(x, a_i, b_i) = (a_i^T x - b_i)^2$$

Logistic Regression

Logistic Loss

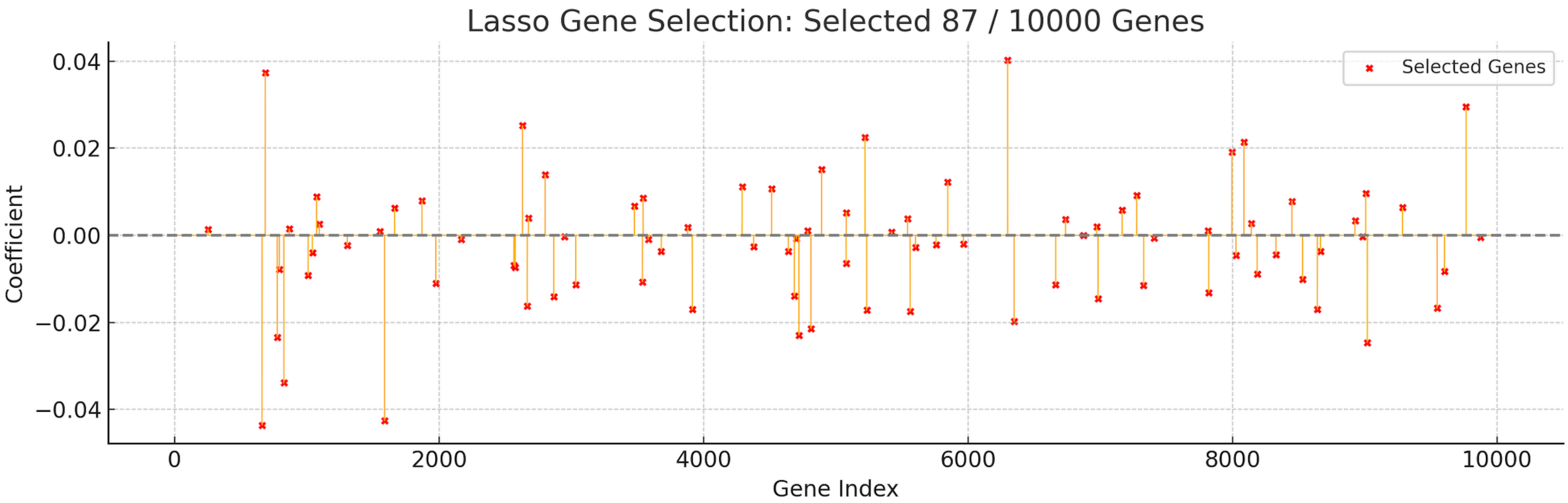
$$l_i(x, a_i, b_i) = \log(1 + \exp(-b_i a_i^T x))$$

Support Vector Machine

Hinge Loss

$$l_i(x, a_i, b_i) = \max\{0, 1 - b_i a_i^T x\}$$

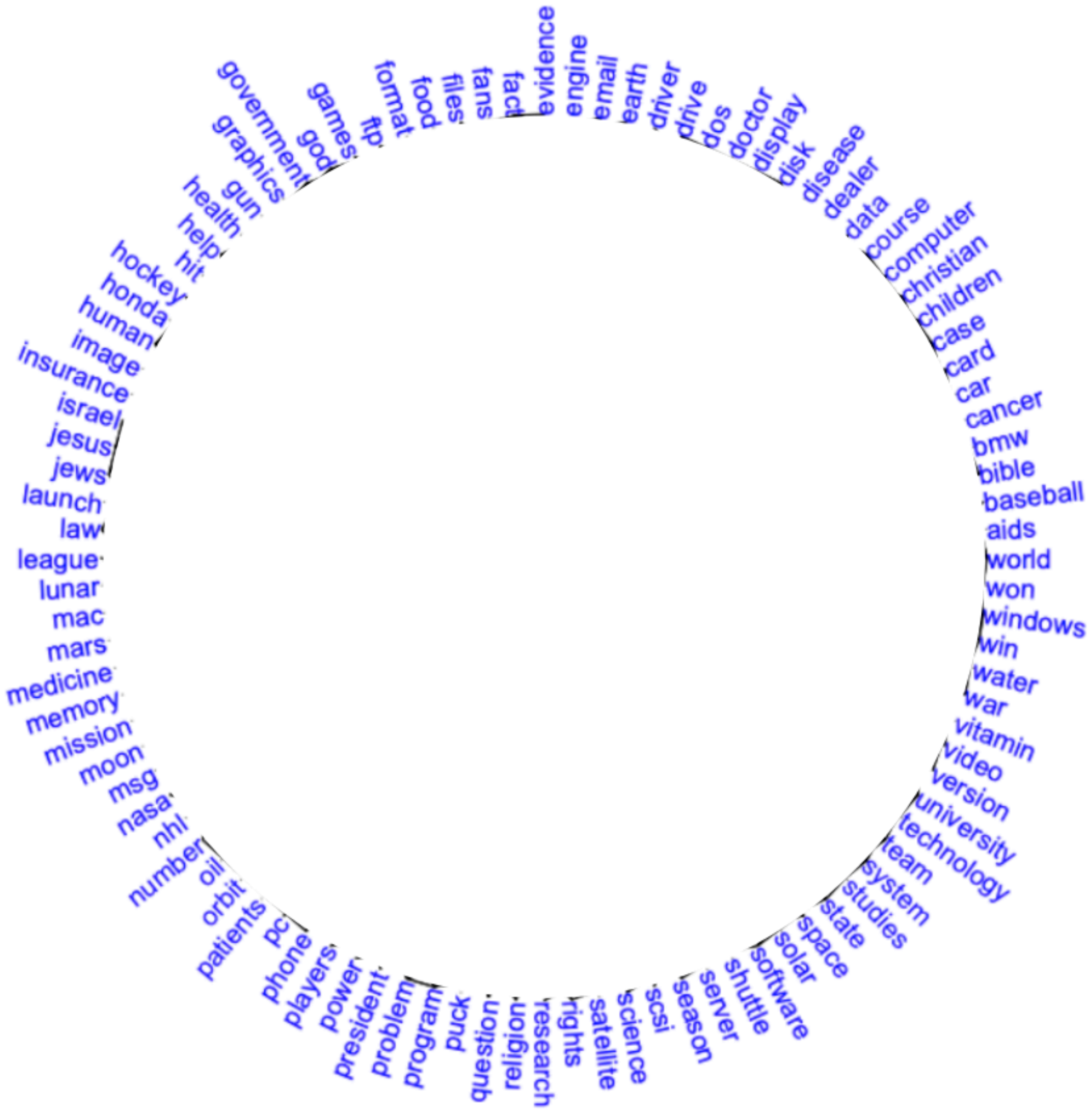
Data Sciences



$$\min_{x \in \mathbb{R}^d} \|Ax - b\|_2^2 + \mu \|x\|_1 \longrightarrow \text{LASSO}$$

Tibshirani R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1996, 58(1): 267-288.

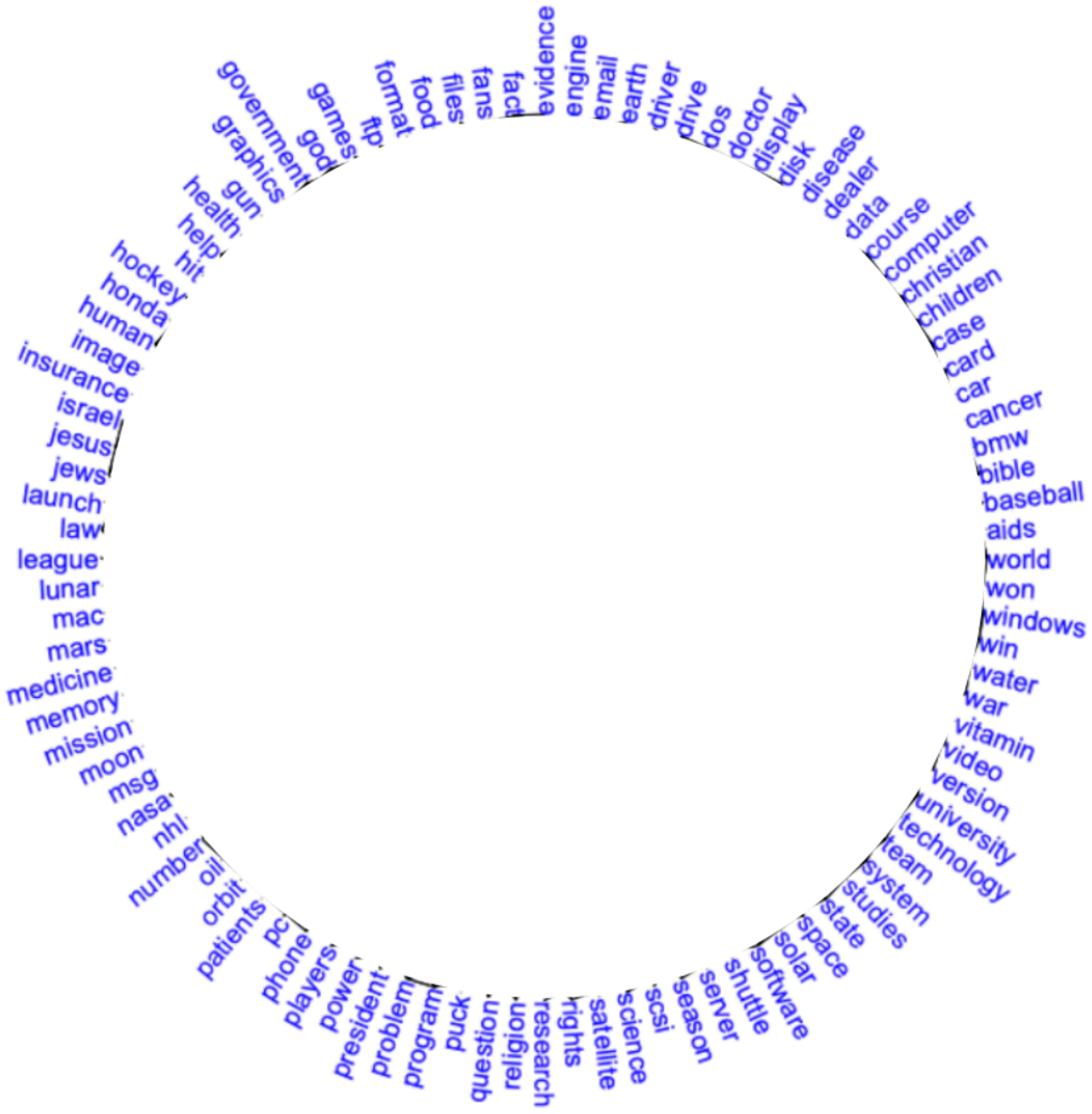
Data Sciences



The data is the publicly available 20newsgroups dataset, which contains binary occurrences of 100 popular words counted from 16, 242 newsgroup postings. On the top level of these postings are 4 main categories: computer, recreation, science and talks. The one-vs-rest scheme for the multi-class classification.

Hua Ouyang et al. “Stochastic alternating direction method of multipliers”. In: International Conference on Machine Learning. 2013, pp. 80–88.

Data Sciences

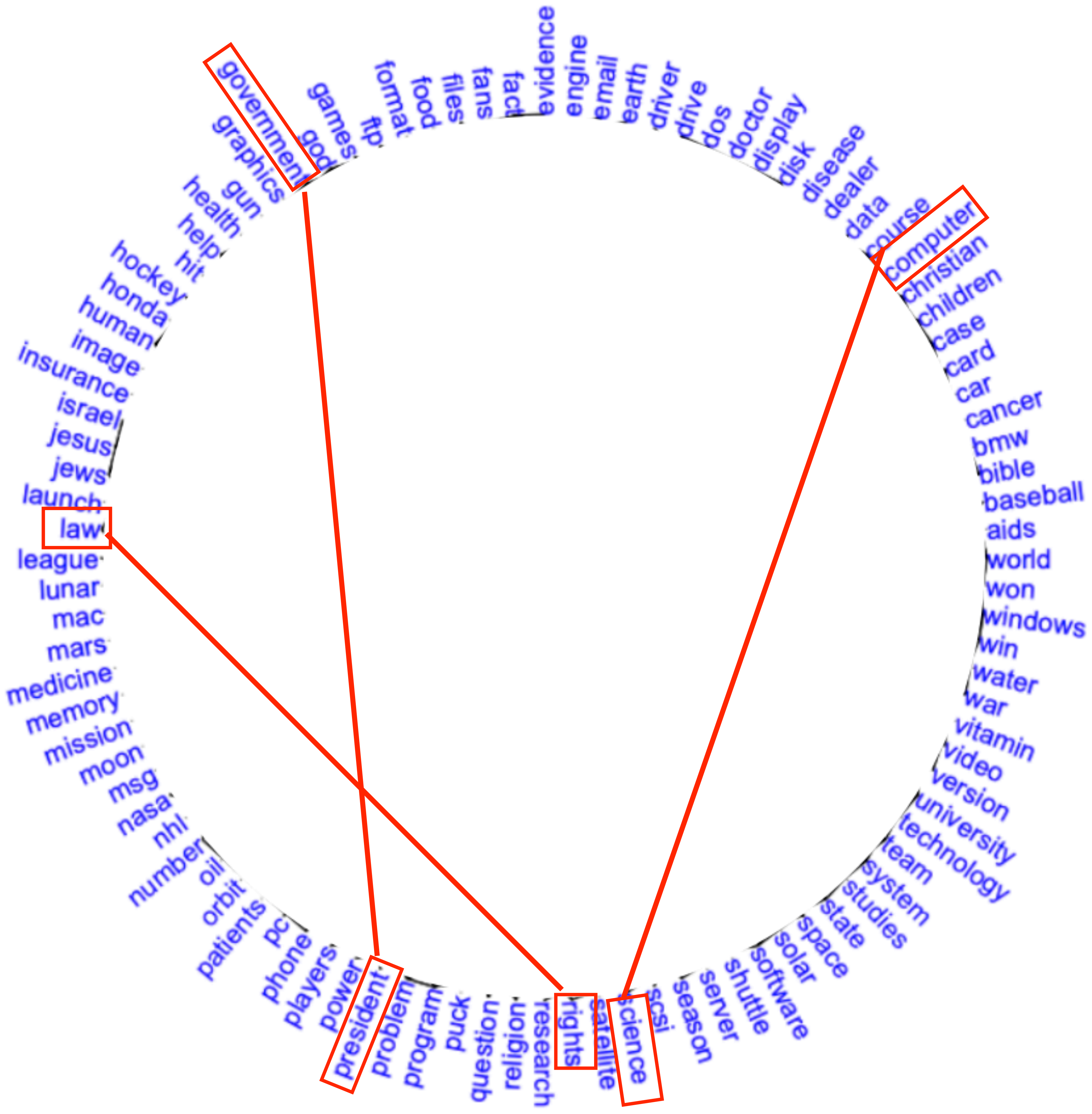


The data is the publicly available 20newsgroups dataset, which contains binary occurrences of 100 popular words counted from 16, 242 newsgroup postings. On the top level of these postings are 4 main categories: computer, recreation, science and talks. The one-vs-rest scheme for the multi-class classification.

$$\min_{x \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \log(1 + \exp(-b_i a_i^T x) + \mu_1 \|x\|_1$$

Hua Ouyang et al. "Stochastic alternating direction method of multipliers". In: International Conference on Machine Learning. 2013, pp. 80–88.

Data Sciences

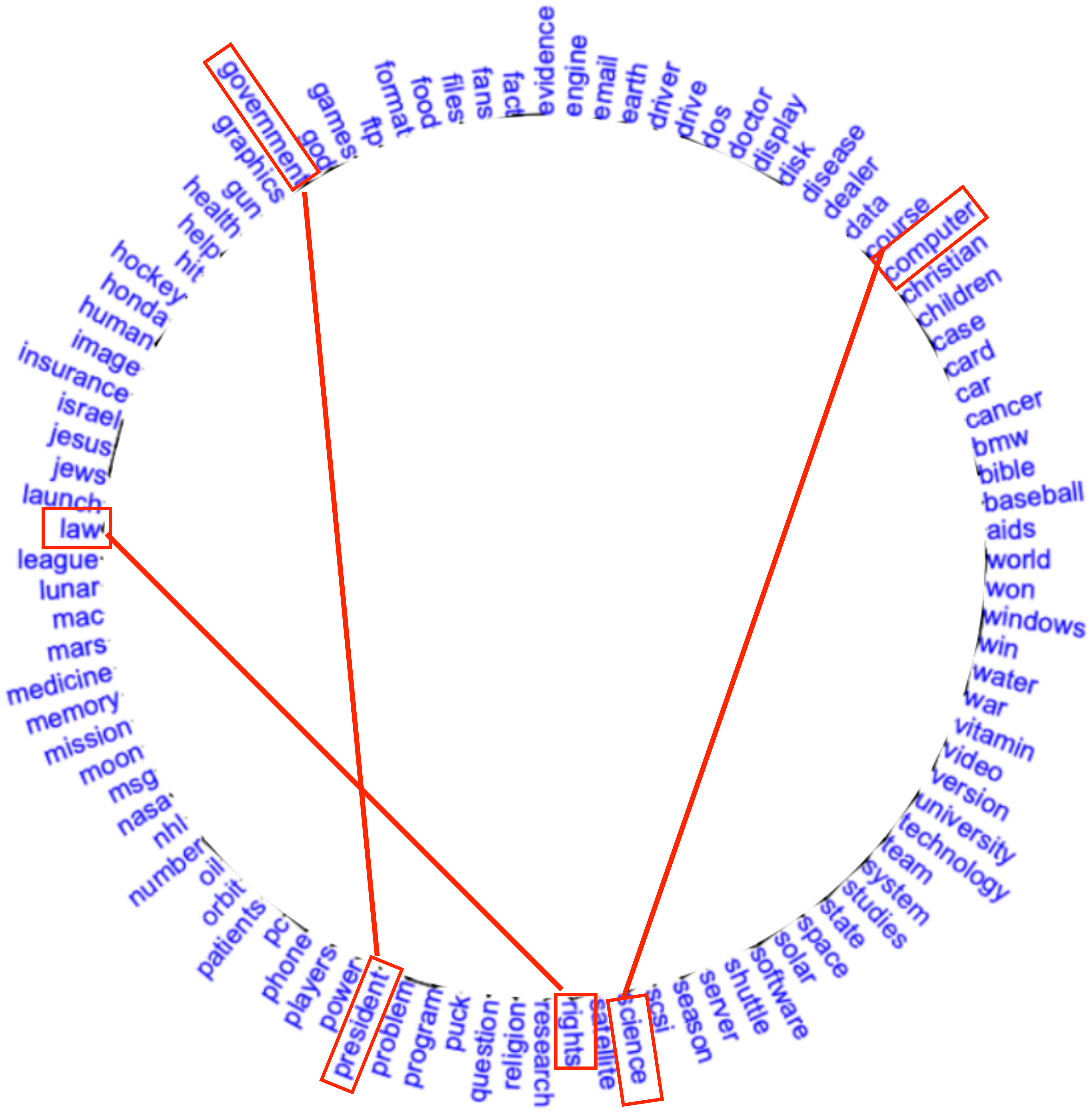


The data is the publicly available 20newsgroups dataset, which contains binary occurrences of 100 popular words counted from 16, 242 newsgroup postings. On the top level of these postings are 4 main categories: computer, recreation, science and talks. The one-vs-rest scheme for the multi-class classification.

$$\min_{x \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \log(1 + \exp(-b_i a_i^T x) + \mu_1 \|x\|_1$$

Hua Ouyang et al. "Stochastic alternating direction method of multipliers". In: International Conference on Machine Learning. 2013, pp. 80–88.

Data Sciences



The data is the publicly available 20newsgroups dataset, which contains binary occurrences of 100 popular words counted from 16, 242 newsgroup postings. On the top level of these postings are 4 main categories: computer, recreation, science and talks. The one-vs-rest scheme for the multi-class classification.

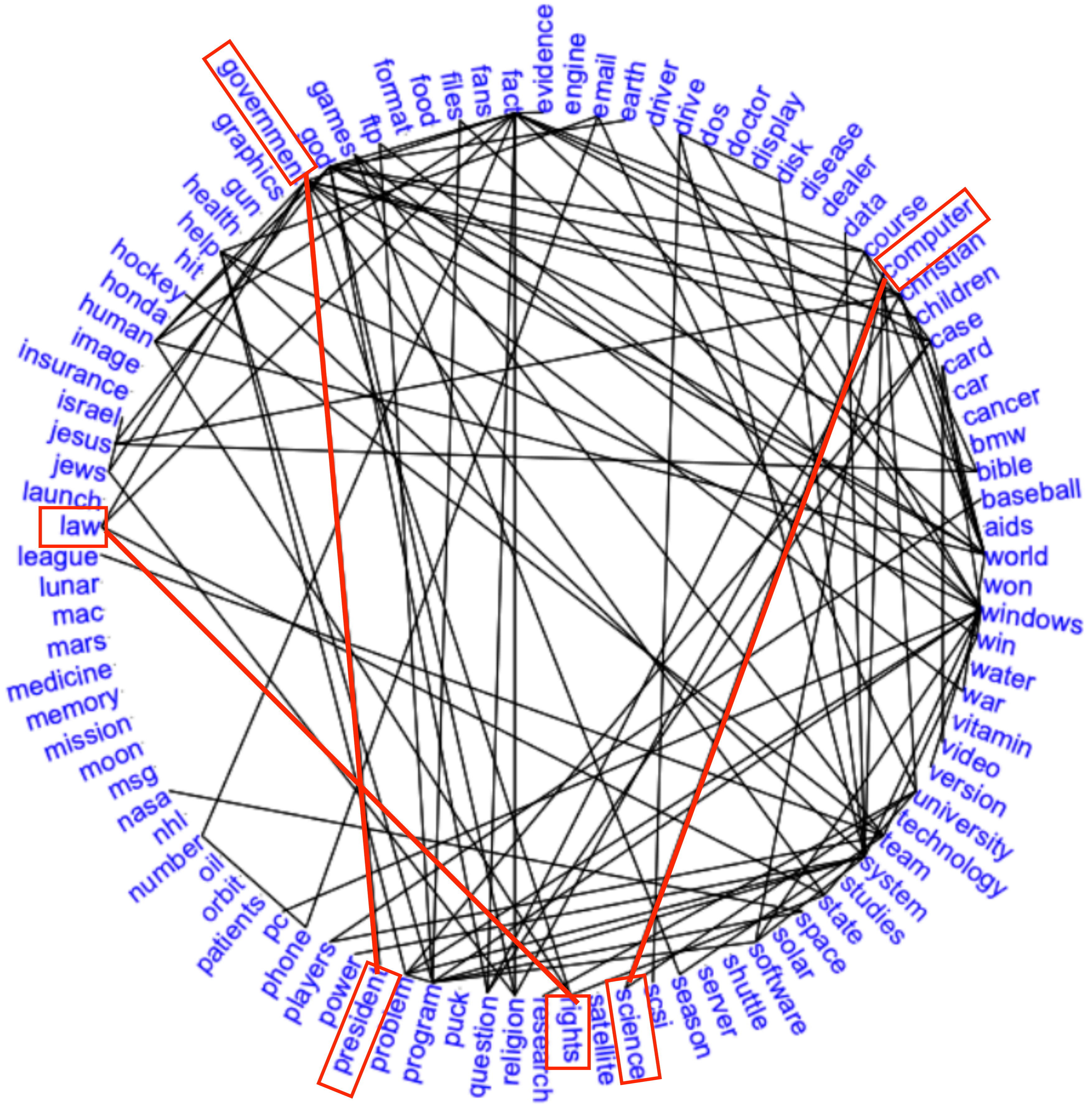
Obtained by sparse inverse covariance selection

$$\min_{x \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \log(1 + \exp(-b_i a_i^T x) + \mu_1 \|x\|_1 + \mu_2 \|Fx\|_1$$

Hua Ouyang et al. “Stochastic alternating direction method of multipliers”. In: International Conference on Machine Learning. 2013, pp. 80–88.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. “Sparse inverse covariance estimation with the graphical lasso”. In: Biostatistics 9.3 (2008), pp. 432–441.

Data Sciences



The data is the publicly available 20newsgroups dataset, which contains binary occurrences of 100 popular words counted from 16, 242 newsgroup postings. On the top level of these postings are 4 main categories: computer, recreation, science and talks. The one-vs-rest scheme for the multi-class classification.

Obtained by sparse inverse covariance selection

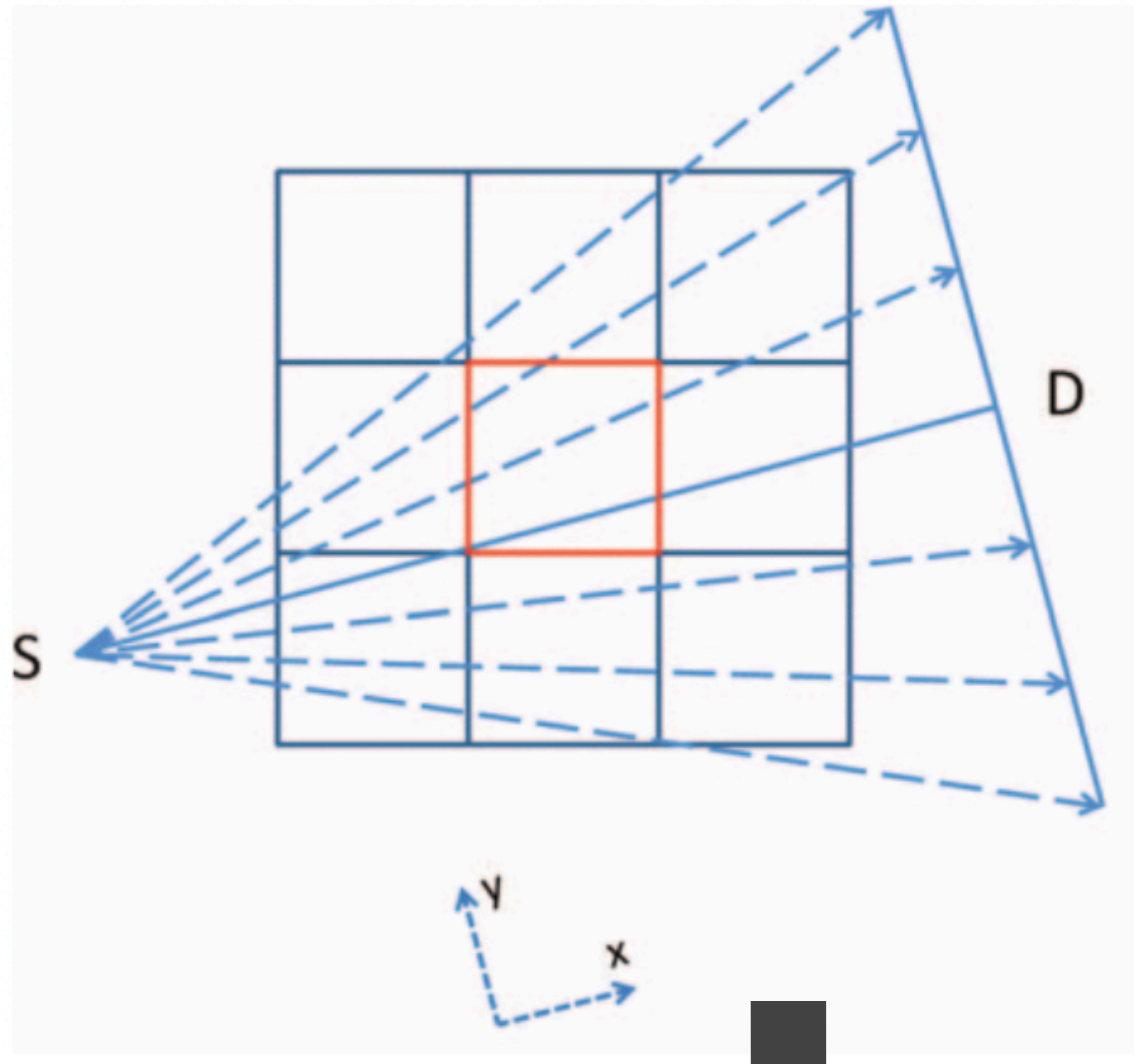
$$\min_{x \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \log(1 + \exp(-b_i a_i^T x) + \mu_1 \|x\|_1 + \mu_2 \|Fx\|_1$$

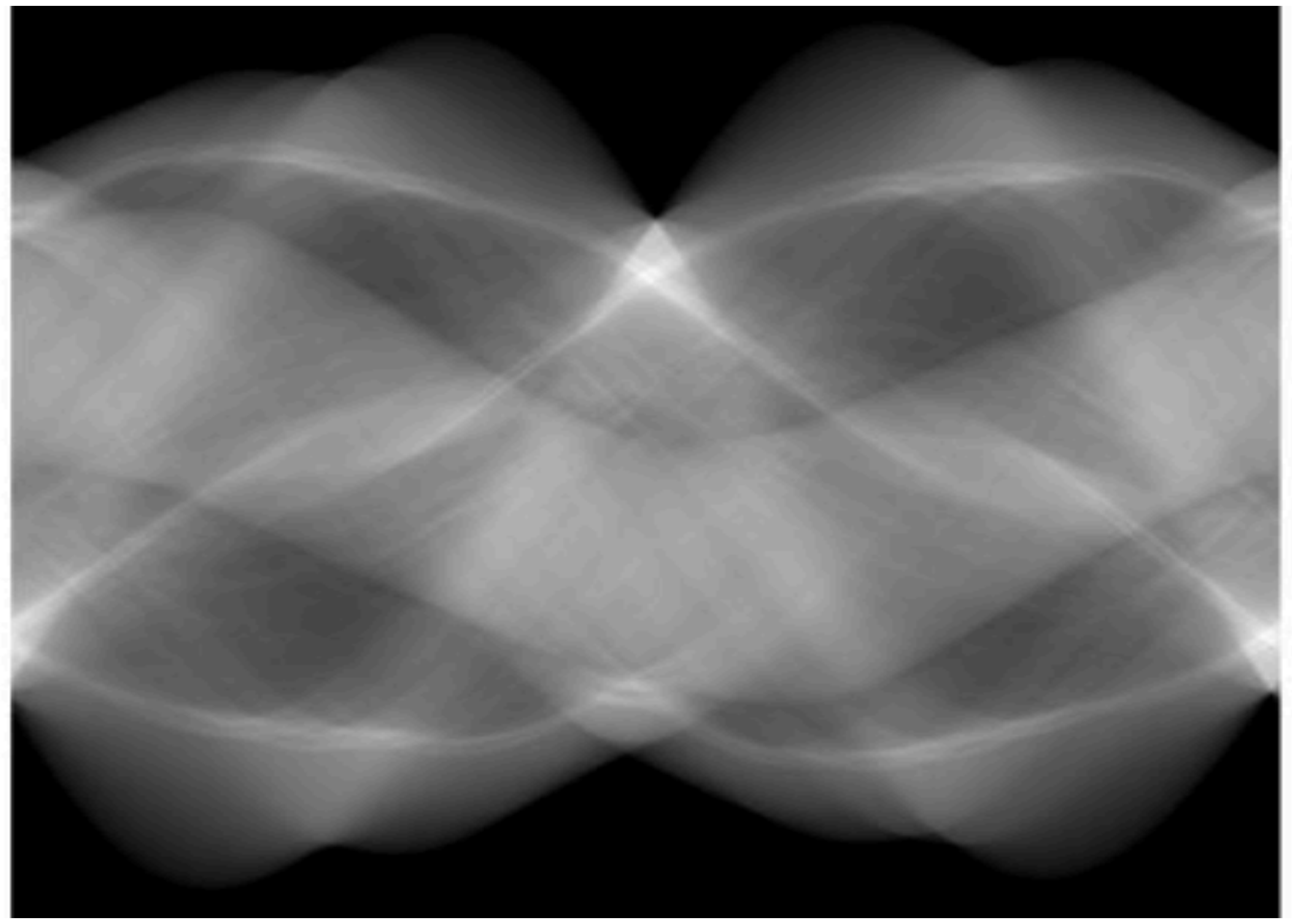
Graph-Guided logistic regression

Hua Ouyang et al. “Stochastic alternating direction method of multipliers”. In: International Conference on Machine Learning. 2013, pp. 80–88.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. “Sparse inverse covariance estimation with the graphical lasso”. In: Biostatistics 9.3 (2008), pp. 432–441.

Computed Tomography (CT) Reconstruction





CT Reconstruction

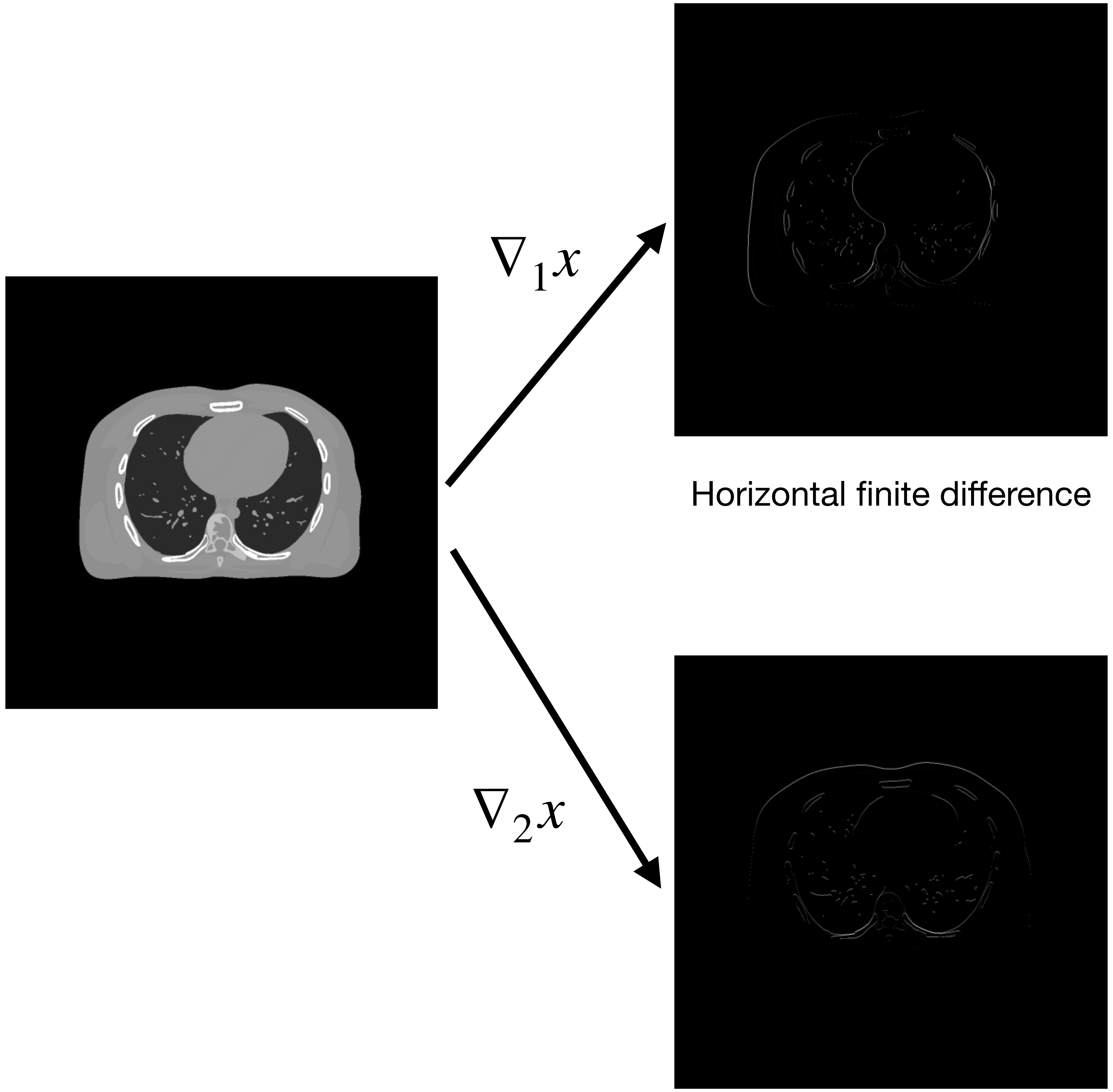
$$\min_{x} \ \frac{1}{2}\|\mathscr{A}x - b\|_2^2 + \mu\|\nabla x\|_{1,2}$$

where

- $\mathscr{A}$ **is X-ray transform and** b **is the observed projection.**

- ∇ **(Isotropic) is discrete gradient operator which aims to recover piecewise constant image.**

- μ **is regularization parameter.**



Horizontal finite difference

Vertical finite difference

Others

Deblur

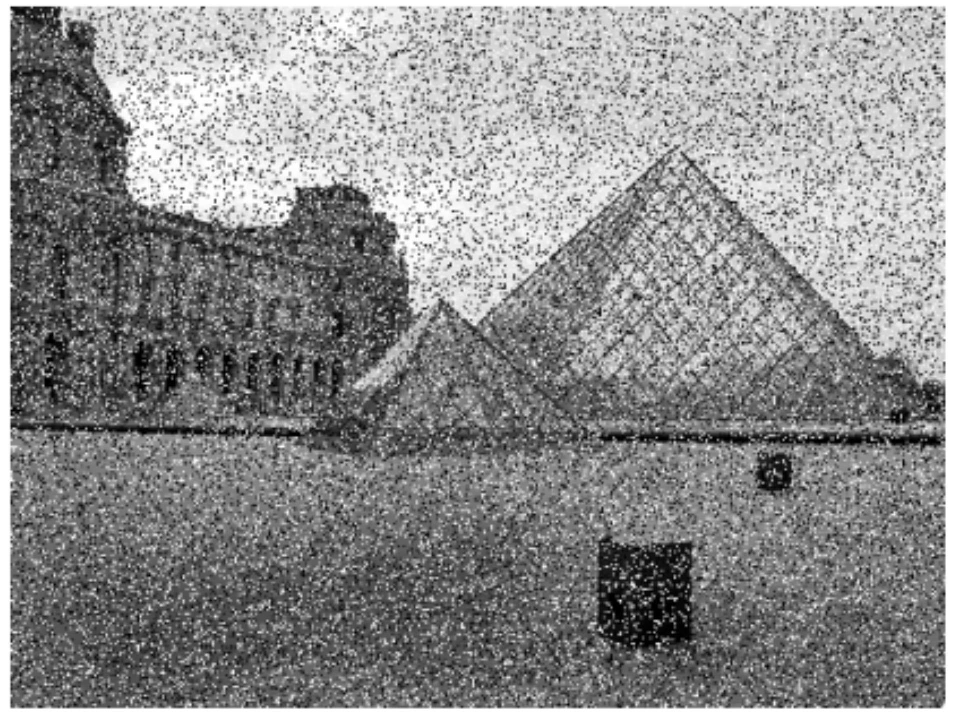
Original

Blur image

Reconstructed

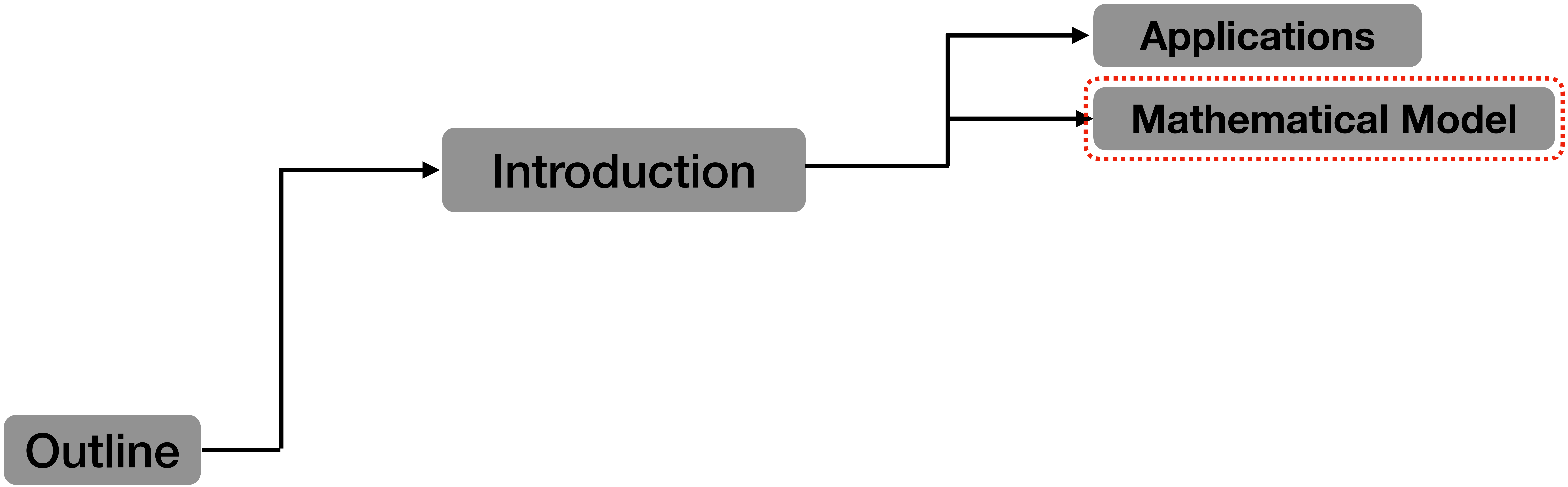
Denoising

Original



Noised Image

Reconstructed


Outline
Introduction
Applications
Mathematical Model

Model

We are devoted to considering the following convex optimization problem

$$\min_{x \in \mathbb{R}^d} \; f(x) + g \circ B(x),$$

where

- $f(x) : \mathbb{R}^d \to]-\infty, +\infty]$ **is a proper convex continuously differentiable with the Lipschitz constant** L_f
- $g(x) : \mathbb{R}^r \to]-\infty, +\infty]$ **is a proper closed convex and may not be differentiable**
- $B \in \mathbb{R}^{r \times d}$ **is a linear mapping**

Model

$$\min_{x\in\mathbb{R}^d} \; f(x) + g \circ B(x),$$

High dimension of the variable

First-order

$$\min_{x\in\mathbb{R}^d} \; f(x) + g \circ B(x),$$

Non-smoothness of the function

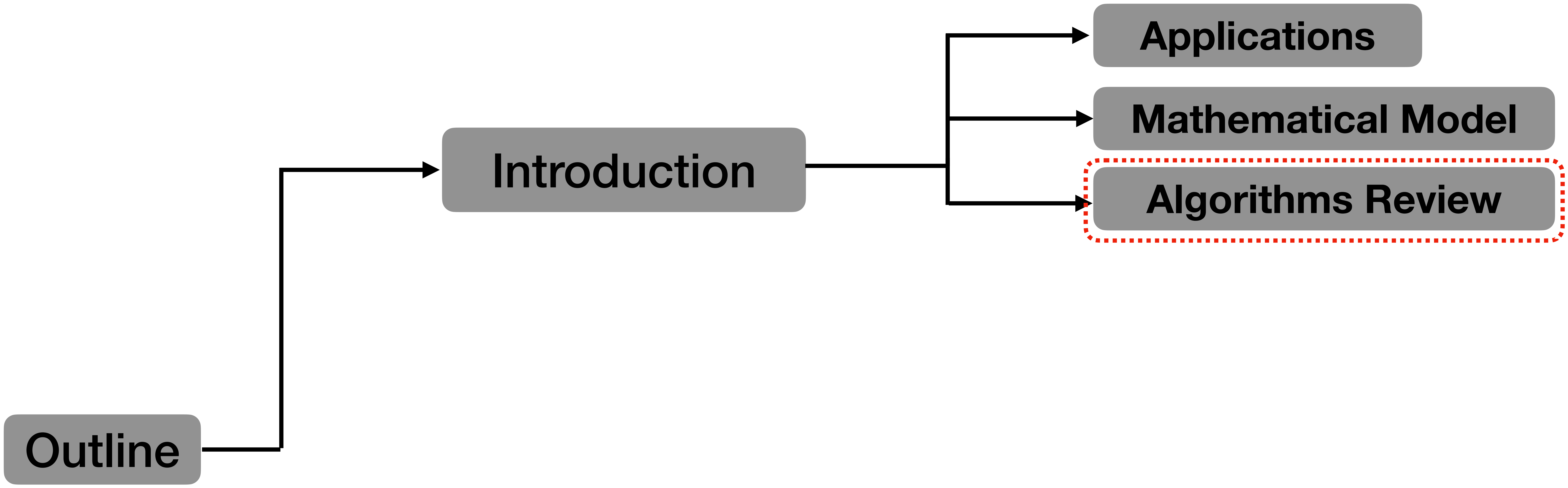
Proximal

$$\min_{x\in\mathbb{R}^d} \; f(x) + g \circ B(x),$$

Splitting structure of the objective

Splitting

Acceleration!


Outline
Introduction
Applications
Mathematical Model
Algorithms Review

Accelerated Proximal Gradient Descent

When $B = I$

$$\min_{x \in \mathbb{R}^d} \; f(x) + g(x)$$

Algorithm: Accelerated Proximal Gradient Descent (FISTA)

Step 1: Set $x_0 = x_1 \in \mathbb{R}^d$ **and choose proper step size** $\gamma > 0$

Step 2: For $k = 1,2,\cdots$

$$z_k = x_k + \frac{k-1}{k+2}\boxed{(x_k - x_{k-1})} \longrightarrow \text{inertial term}$$

$$\boxed{\begin{aligned} x_{k+\frac{1}{2}} &= z_k - \gamma \nabla f(z_k) \\ x_{k+1} &= \mathrm{Prox}_{\gamma g}(x_{k+\frac{1}{2}}) \end{aligned}} \longrightarrow \text{PGD}$$

until the stop criterion is satisfied

Amir Beck and Marc Teboulle. "A Fast Iterative Shrinkage-Thresholding Algorithm for linear inverse problems". In: SIAM Journal on imaging sciences 2.1 (2009), pp. 183–202.

Nesterov Accelerated Gradient (NAG)

When $B = I$

$$\min_{x \in \mathbb{R}^d} \; f(x) + g(x)$$

Algorithm: Nesterov Accelerated Gradient (NAG)
Step 1: Choose $x_0 = x_0^{\mathrm{ag}} \in \mathbb{R}^d$ **and proper step size** γ**.**
Step 2: For $k = 1, 2, \cdots$
$\theta_k = 2/(k+1)$
$\tilde{x}_k = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_k$
$x_{k+1} = \mathrm{Prox}_{(\gamma/\theta_k)g}(x_k - \gamma/\theta_k \nabla f(\tilde{x}_k))$
$x_{k+1}^{\mathrm{ag}} = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_{k+1}$
until the stop criterion is satisfied.

Y.Nesterov. “On an approach to the construction of optimal methods of minimization of smooth convex functions”. Ekonom. i. Mat. Metody 24 (1988),509-517

Tseng P. “On accelerated proximal gradient methods for convex-concave optimization”. SIAM Journal on Optimization, 2008, 2(3).

$B \neq I$

When $B \neq I$, proximal gradient have to solve

$$\mathrm{Prox}_{\gamma g \circ B}(z) = \arg\min_{u \in \mathbb{R}^d} \{\gamma g \circ B(u) + \frac{1}{2}\|u - z\|_2^2\}$$

which is as difficult as original problem

$$\min_{x \in \mathbb{R}^d} \ f(x) + g \circ B(x),$$

- **Alternating Direction Method of Multipliers (ADMM)**
- **Modified Primal Dual Hybrid Gradient (PDHGm or Chambolle-Pock (CP))**
- **Primal Dual Fixed Point (PDFP)**

ADMM

ADMM

$$\min_{x \in \mathbb{R}^d} \; f(x) + g \circ B(x)$$

$$\begin{aligned} \min_{x \in \mathbb{R}^d, z \in \mathbb{R}^r} \quad & f(x) + g(z) \\ s.t. \quad & Bx = z \end{aligned}$$

$$\mathcal{L}(x, z, y) = f(x) + g(z) + \langle y, Bx - z \rangle + \frac{\rho_k}{2} \| Bx - z \|_2^2$$

Augmented Lagrangian

ADMM

$$\mathscr{L}(x,z,y) = f(x) + g(z) + \langle y, Bx - z\rangle + \frac{\rho_k}{2}\|Bx - z\|_2^2$$

Algorithm: Alternating Direction method of Multipliers (ADMM)

Step 1 : Set $x_1 \in \mathbb{R}^d\ z_1, y_1 \in \mathbb{R}^r$ **and choose proper** $\rho_k > 0$.

Step 2 : For $k = 1,2,\cdots$

$$x_{k+1} = \arg\min_{x\in\mathbb{R}^d} f(x) + \langle y_k, Bx - z_k\rangle + \frac{\rho_k}{2}\|Bx - z_k\|_2^2$$

$$z_{k+1} = \arg\min_{z\in\mathbb{R}^r} g(z) + \langle y_k, Bx_{k+1} - z\rangle + \frac{\rho_k}{2}\|Bx_{k+1} - z\|_2^2$$

$$y_{k+1} = y_k + \rho_k(Bx_{k+1} - z_{k+1})$$

until the stop criterion is satisfied.

D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite-element approximations, Comput. Math. Appl., 2 (1976), pp. 17–40.

Goldstein T, Osher S. The split Bregman method for L1-regularized problems[J]. SIAM journal on imaging sciences, 2009, 2(2): 323-343.

ADMM

Algorithm: Alternating Direction method of Multipliers (ADMM)

Step 1 : Set $x_1 \in \mathbb{R}^d$ $z_1, y_1 \in \mathbb{R}^r$ **and choose proper** $\rho_k > 0$.
Step 2 : For $k = 1,2,\cdots$

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^d} f(x) + \langle y_k, Bx - z_k \rangle + \frac{\rho_k}{2}\|Bx - z_k\|_2^2$$

$$z_{k+1} = \arg\min_{z \in \mathbb{R}^r} g(z) + \langle y_k, Bx_{k+1} - z \rangle + \frac{\rho_k}{2}\|Bx_{k+1} - z\|_2^2$$

$$y_{k+1} = y_k + \rho_k(Bx_{k+1} - z_{k+1})$$

until the stop criterion is satisfied.

The second subproblem of ADMM

$$z_{k+1} = \arg\min_{z \in \mathbb{R}^r} g(z) + \langle y_k, Bx_{k+1} - z \rangle + \frac{\rho_k}{2}\|Bx_{k+1} - z\|_2^2$$

$$\Downarrow$$

$$z_{k+1} = \arg\min_{z \in \mathbb{R}^r} g(z) + \frac{\rho_k}{2}\|Bx_{k+1} + y_k/\rho_k - z\|_2^2$$

$$\Downarrow$$

$$z_{k+1} = \mathrm{Prox}_{\frac{g}{\rho_k}}(Bx_{k+1} + y_k/\rho_k)$$

$$\Downarrow$$

Usually admits a closed-form solution

ADMM

Algorithm: Alternating Direction method of Multipliers (ADMM)

Step 1 : Set $x_1 \in \mathbb{R}^d$ $z_1, y_1 \in \mathbb{R}^m$ **and choose proper** $\rho_k > 0$.
Step 2 : For $k = 1,2,\cdots$

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^d} f(x) + \langle y_k, Bx - z_k \rangle + \frac{\rho_k}{2}\|Bx - z_k\|_2^2$$

$$z_{k+1} = \arg\min_{z \in \mathbb{R}^r} g(z) + \langle y_k, Bx_{k+1} - z \rangle + \frac{\rho}{2}\|Bx_{k+1} - z\|_2^2$$

$$y_{k+1} = y_k + \rho(Bx_{k+1} - z_{k+1})$$

until the stop criterion is satisfied.

The first subproblem of ADMM

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^d} f(x) + \langle y_k, Bx - z_k \rangle + \frac{\rho_k}{2}\|Bx - z_k\|_2^2$$

$$\Downarrow$$

Solving this problem is generally not easy

$$\Downarrow$$

Solve it by using other algorithms,
e.g. (Proximal) gradient descent or its accelerations

Linearized Preconditioned ADMM (LP-ADMM)

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^d} f(x) + \langle y_k, Bx - z_k \rangle + \frac{\rho_k}{2}\|Bx - z_k\|_2^2$$

$$\Downarrow$$

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^d} f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \langle y_k, Bx - z_k \rangle + \frac{\rho_k}{2}\|Bx_k - z_k\|_2^2 + \rho_k \langle B^T(Bx_k - z_k), x - x_k \rangle + \frac{1}{2\gamma_k}\|x - x_k\|_2^2$$

$$\Downarrow$$

Find Some point x_{k+1} such that $\nabla f(x_k) + B^T y_k + \rho_k B^T(Bx_k - z_k) + \frac{1}{\gamma_k}(x_{k+1} - x_k) = 0$

$$\Downarrow$$

$$x_{k+1} = x_k - \gamma_k(\nabla f(x_k) + B^T y_k + \rho_k B^T(Bx_k - z_k))$$

$$\Downarrow$$

One step gradient descent of Augmented Lagrangian

LP-ADMM

$$\min_{x \in \mathbb{R}^d, y \in \mathbb{R}^r} \quad f(x) + g(z)$$
$$s.t. \quad Bx = z$$

Algorithm: Linearized Preconditioned ADMM (LP-ADMM)

Step 1 : Set $x_1, z_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $C > 0,$
let $\rho_k := \rho = C/\|B\|, \gamma_k := \gamma = 1/(L_f + \rho\|B\|^2)$**.**
Step 2 : For $k = 1,2,\cdots$

$$x_{k+1} = x_k - \gamma_k(\nabla f(x_k) + \rho_k B^T(Bx_k - z_k) + B^T y_k)$$

$$z_{k+1} = \text{Prox}_{g/\rho}(Bx_{k+1} + y_k/\rho_k)$$

$$y_{k+1} = y_k + \rho(Bx_{k+1} - z_{k+1})$$

until the stop criterion is satisfied.

Ouyang Y, Chen Y, Lan G, et al. An accelerated linearized alternating direction method of multipliers[J]. SIAM Journal on Imaging Sciences, 2015, 8(1): 644-681.

LP-ADMM

$$\min_{x\in\mathbb{R}^d, z\in\mathbb{R}^r} \quad f(x) + g(z)$$
$$s.t. \quad Bx = z$$

Algorithm: Linearized Preconditioned ADMM (LP-ADMM)

Step 1 : Set $x_1, z_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $\rho_k := \rho = C/\|B\|, \gamma_k := \gamma = 1/(L_f + \rho\|B\|^2)$.

Step 2 : For $k = 1,2,\cdots$

$$\boxed{x_{k+1} = x_k - \gamma_k(\nabla f(x_k) + \rho_k B^T(Bx_k - z_k) + B^T y_k)}$$

$$z_{k+1} = \mathrm{Prox}_{g/\rho_k}(Bx_{k+1} + y_k/\rho_k)$$

$$y_{k+1} = y_k + \rho_k(Bx_{k+1} - z_{k+1})$$

until the stop criterion is satisfied.

$$\bar{x}_k = \frac{1}{k}\sum_{i=1}^{k} x_i$$

$$f(\bar{x}_k) + g \circ B(\bar{x}_k) - (f^* + g^*) \le \mathcal{O}\left(\frac{L_f}{k}\right) + \mathcal{O}\left(\frac{\|B\|}{k}\right)$$

L_f is usually much larger than $\|B\|$

Using Nesterov acceleration technique to improve the rate in terms of its dependence on L_f

Ouyang Y, Chen Y, Lan G, et al. An accelerated linearized alternating direction method of multipliers[J]. SIAM Journal on Imaging Sciences, 2015, 8(1): 644-681.

AADMM

Algorithm: Accelerated ADMM (AADMM)

Step 1 : Set $x_1, z_1, \tilde{x}_1, \bar{x}_1, \in \mathbb{R}^d, p_1 \in \mathbb{R}^m$ **and choose proper** $\rho > 0$, $\sigma_k = (k-1)\rho/k$, $\gamma_k = k/(2/L_f + \rho k\|B\|^2)$

Step 2 : For $k = 1,2,\cdots$

$$\theta_k = 2/(k+1)$$
$$\tilde{x}_k = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_k$$
$$x_{k+1} = x_k - \gamma_k(\nabla f(\tilde{x}_k) + \sigma_k B^T(Bx_k - z_k) + B^T y_k)$$
$$x_{k+1}^{\mathrm{ag}} = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k z_{k+1}$$
$$z_{k+1} = \mathrm{Prox}_{g/\rho}(Bx_{k+1} + y_k/\rho)$$
$$y_{k+1} = y_k + \rho(Bx_{k+1} - z_{k+1})$$

until the stop criterion is satisfied.

$$f(x_k^{\mathrm{ag}}) + g \circ B(x_k^{\mathrm{ag}}) - (f^* + g^*) \leq \mathcal{O}\left(\frac{L_f}{k^2}\right) + \mathcal{O}\left(\frac{\|B\|}{k}\right)$$

Ouyang Y, Chen Y, Lan G, et al. An accelerated linearized alternating direction method of multipliers[J]. SIAM Journal on Imaging Sciences, 2015, 8(1): 644-681.

PDHGm/CP

APD

$$\min_{x \in \mathbb{R}^d} \; f(x) + g \circ B(x)$$

$$\min_{x \in \mathbb{R}^d} \max_{y \in \mathbb{R}^r} \; f(x) + \langle Bx, y \rangle - g^*(y)$$

$$g^*(y) = \max_{u \in \mathbb{R}^r} \; \langle y, u \rangle - g(u)$$

Conjugate function of g

PDHGm (a.k.a CP)

$$\min_{x\in\mathbb{R}^d}\max_{y\in\mathbb{R}^r} f(x) + \langle Bx, y\rangle - g^*(y)$$

Algorithm: Modified Primal Dual Hybrid Gradient (PDHGm)

Step 1 : Set $x_1, \bar{x}_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $\sigma, \tau > 0$ **such that** $\sigma\tau\|B\|^2 < 1$ **and** $\alpha \in (0,1]$.

Step 2 : For $k = 1,2,\cdots$

$$y_{k+1} = \mathrm{Prox}_{\sigma g^*}(\sigma B\bar{x}_k + y_k)$$
$$x_{k+1} = \mathrm{Prox}_{\tau f}(x_k - \tau B^T y_{k+1})$$
$$\bar{x}_{k+1} = x_{k+1} + \alpha_k(x_{k+1} - x_k)$$

until the stop criterion is satisfied.

Antonin Chambolle and Thomas Pock. “A first-order primal-dual algorithm for convex problems with applications to imaging”. In: Journal of mathematical imaging and vision 40.1 (2011), pp. 120–145.
Ernie Esser, Xiaoqun Zhang, and Tony F Chan. “A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science”. SIAM Journal on Imaging Sciences 3.4 (2010), pp. 1015–1046.

PDHGm/CP

The second subproblem of PDHGm

$$x_{k+1} = \mathrm{Prox}_{\tau f}(x_k - \tau B^T y_{k+1})$$

$$x_{k+1} = \arg\min_x \boxed{f(x) + \frac{1}{2\tau}\|x - (x_k - \tau B^T y_{k+1})\|_2^2}$$

Linearization

$$x_{k+1} = \arg\min_{x\in\mathbb{R}^r} f(x_k) + \langle \nabla f(x_k), x - x_k\rangle + \frac{1}{2\eta_k}\|x - x_k\|_2^2 + \langle x, B^T y_{k+1}\rangle + \frac{1}{2\tau}\|x - x_k\|_2^2$$

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k) - \gamma_k B^T y_{k+1}$$

One step proximal gradient descent

Linearized PDHGm (LPDHGm)

Algorithm: Linearized Primal Dual Hybrid Gradient (LPDHGm)

Step 1 : Set $x_1, \bar{x}_1, \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $0 < \alpha_k = \sigma_{k-1}/\sigma_k = \gamma_{k-1}/\gamma_k \leq 1,\ L_f\sigma_k + \|B\|^2\gamma_k\sigma_k \leq 1,$ **.**

Step 2 : For $k = 1,2,\cdots$

$$y_{k+1} = \mathrm{Prox}_{\sigma_k g^*}(y_k + \sigma_k B\bar{x}_k)$$

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k) - \gamma_k B^T y_{k+1}$$

$$\bar{x}_{k+1} = x_{k+1} + \alpha_k(x_{k+1} - x_k)$$

until the stop criterion is satisfied.

Chen Y, Lan G, Ouyang Y. Optimal primal-dual methods for a class of saddle point problems[J]. SIAM Journal on Optimization, 2014, 24(4): 1779-1814.

LPDHGm

$$\min_{x \in \mathbb{R}^d} \max_{y \in \mathbb{R}^r} \; f(x) + \langle Bx, y \rangle - g^*(y)$$

Algorithm: Linearized Primal Dual Hybrid Gradient (LPDHGm)
Step 1 : Set $x_1, \bar{x}_1, \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $0 < \alpha_k = \sigma_{k-1}/\sigma_k = \gamma_{k-1}/\gamma_k \le 1,\; L_f\sigma_k + \|B\|^2\gamma_k\sigma_k \le 1,$.
Step 2 : For $k = 1,2,\cdots$
$y_{k+1} = \text{Prox}_{\sigma_k g^*}(y_k + \sigma_k B\bar{x}_k)$
$\boxed{x_{k+1} = x_k - \gamma_k \nabla f(x_k) - \gamma_k B^T y_{k+1}}$
$\bar{x}_{k+1} = x_{k+1} + \alpha_k(x_{k+1} - x_k)$
until the stop criterion is satisfied.

$$Q(\tilde{x}, \tilde{y}, x, y) = f(\tilde{x}) + \langle B\tilde{x}, y \rangle - g^*(y)$$
$$-(f(x) + \langle Bx, \tilde{y} \rangle - g^*(\tilde{y}))$$

$$\mathcal{G}(\tilde{x}, \tilde{y}) = \sup_{x \in B_1, y \in B_2} Q(\tilde{x}, \tilde{p}, x, y)$$

Partial primal dual gap

$$\mathcal{G}(\bar{x}_k, \bar{y}_k) \le \mathcal{O}\left(\boxed{\frac{L_f}{k}}\right) + \mathcal{O}\left(\frac{\|B\|}{k}\right)$$

$$\bar{x}_k = \frac{1}{k}\sum_{i=1}^{k} x_i \qquad \bar{y}_k = \frac{1}{k}\sum_{i=1}^{k} y_i,$$

Chen Y, Lan G, Ouyang Y. Optimal primal-dual methods for a class of saddle point problems[J]. SIAM Journal on Optimization, 2014, 24(4): 1779-1814.

Accelerated Linearized PDHGm (ALPDHGm/APD)

Algorithm: APD

Step 1 : Set $x_1 = x_1^{\mathrm{ag}}, z_1, \bar{x}_1, \in \mathbb{R}^d, y_1 \in \mathbb{R}^m$ **and choose proper**
$C > 0$**,** $\tau_k = C/\|B\|$**,** $\gamma_k = k/(2L_f + k\|B\|C), \alpha_{k+1} = k/(k+1)$

Step 2 : For $k = 1,2,\cdots$

$$\theta_k = 2/(k+1), \alpha_k = (k-1)/k$$

$$y_{k+1} = \mathrm{Prox}_{\tau_k g^*}(y_k + \tau_k B\bar{x}_k)$$

$$\tilde{x}_k = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_k$$

$$x_{k+1} = x_k - \gamma_k \nabla f(\tilde{x}_k) - \gamma_k B^T y_{k+1}$$

$$x_{k+1}^{\mathrm{ag}} = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_{k+1}$$

$$\bar{x}_{k+1} = x_{k+1} + \alpha_{k+1}(x_{k+1} - x_k)$$

until the stop criterion is satisfied.

Chen Y, Lan G, Ouyang Y. Optimal primal-dual methods for a class of saddle point problems[J]. SIAM Journal on Optimization, 2014, 24(4): 1779-1814.

APD

Algorithm: APD

Step 1 : Set $x_1 = x_1^{\mathrm{ag}}, z_1, \bar{x}_1, \in \mathbb{R}^d, y_1 \in \mathbb{R}^m$ **and choose proper** $C > 0$, $\tau_k = C/\|B\|$, $\gamma_k = k/(2L_f + k\|B\|C)$

Step 2 : For $k = 1,2,\cdots$

$$\theta_k = 2/(k+1), \alpha_k = (k-1)/k$$

$$y_{k+1} = \mathrm{Prox}_{\tau_k g^*}(y_k + \tau_k B \bar{x}_k)$$

$$\tilde{x}_k = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_k$$

$$x_{k+1} = x_k - \gamma_k \nabla f(\tilde{x}_k) - \gamma_k B^T y_{k+1}$$

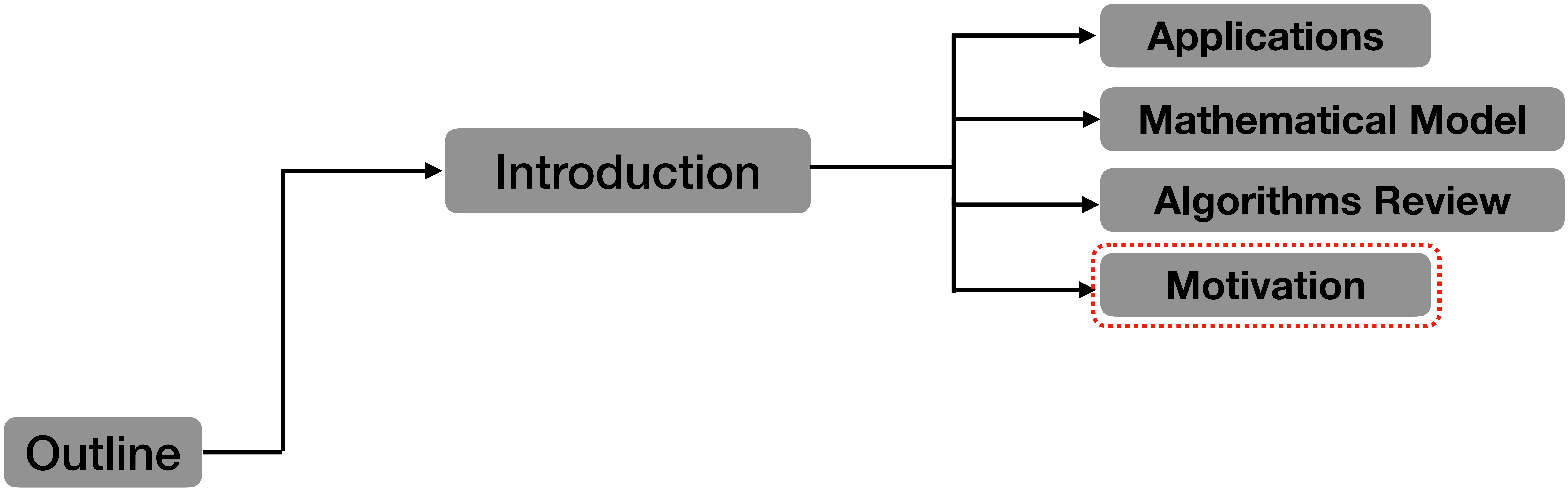
$$x_{k+1}^{\mathrm{ag}} = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_{k+1}$$

$$\bar{x}_{k+1} = x_{k+1} + \alpha_k(x_{k+1} - x_k)$$

until the stop criterion is satisfied.

$$\mathcal{G}(x_{k+1}^{\mathrm{ag}}, y_{k+1}^{\mathrm{ag}}) \leq \mathcal{O}\left(\frac{L_f}{k^2}\right) + \mathcal{O}\left(\frac{\|B\|}{k}\right)$$

Chen Y, Lan G, Ouyang Y. Optimal primal-dual methods for a class of saddle point problems[J]. SIAM Journal on Optimization, 2014, 24(4): 1779-1814.


Outline
Introduction
Applications
Mathematical Model
Algorithms Review
Motivation

$$\min_{x\in\mathbb{R}^d} \; f(x) + g \circ B(x)$$

Algorithm: Accelerated ADMM (AADMM)

Step 1 : **Set** $x_1, z_1, \tilde{x}_1, \bar{x}_1, \in \mathbb{R}^d, p_1 \in \mathbb{R}^m$ **and choose proper** $\rho > 0$, $\sigma_k = (k-1)\rho/k$, $\gamma_k = k/(2/L_f + \rho k\|B\|^2)$

Step 2 : **For** $k = 1,2,\cdots$

$\theta_k = 2/(k+1)$

$\tilde{x}_k = (1-\theta_k)x_k^{ag} + \theta_k x_k$

$x_{k+1} = x_k - \gamma_k(\nabla f(\tilde{x}_k) + \sigma_k B^T(Bx_k - z_k) + B^T y_k)$

$x_{k+1}^{ag} = (1-\theta_k)x_k^{ag} + \theta_k z_{k+1}$

$z_{k+1} = \text{Prox}_{g/\rho}(Bx_{k+1} + y_k/\rho)$

$y_{k+1} = y_k + \rho(Bx_{k+1} - z_{k+1})$

until the stop criterion is satisfied.

Algorithm: APD

Step 1 : **Set** $x_1 = x_1^{ag}, z_1, \bar{x}_1, \in \mathbb{R}^d, y_1 \in \mathbb{R}^m$ **and choose proper** $C > 0$, $\tau_k = C/\|B\|$, $\gamma_k = k/(2L_f + k\|B\|C)$, $\alpha_{k+1} = k/(k+1)$

Step 2 : **For** $k = 1,2,\cdots$

$\theta_k = 2/(k+1), \alpha_k = (k-1)/k$

$y_{k+1} = \text{Prox}_{\tau_k g^*}(y_k + \tau_k B\bar{x}_k)$

$\tilde{x}_k = (1-\theta_k)x_k^{ag} + \theta_k x_k$

$x_{k+1} = x_k - \gamma_k \nabla f(\tilde{x}_k) - \gamma_k B^T y_{k+1}$

$x_{k+1}^{ag} = (1-\theta_k)x_k^{ag} + \theta_k x_{k+1}$

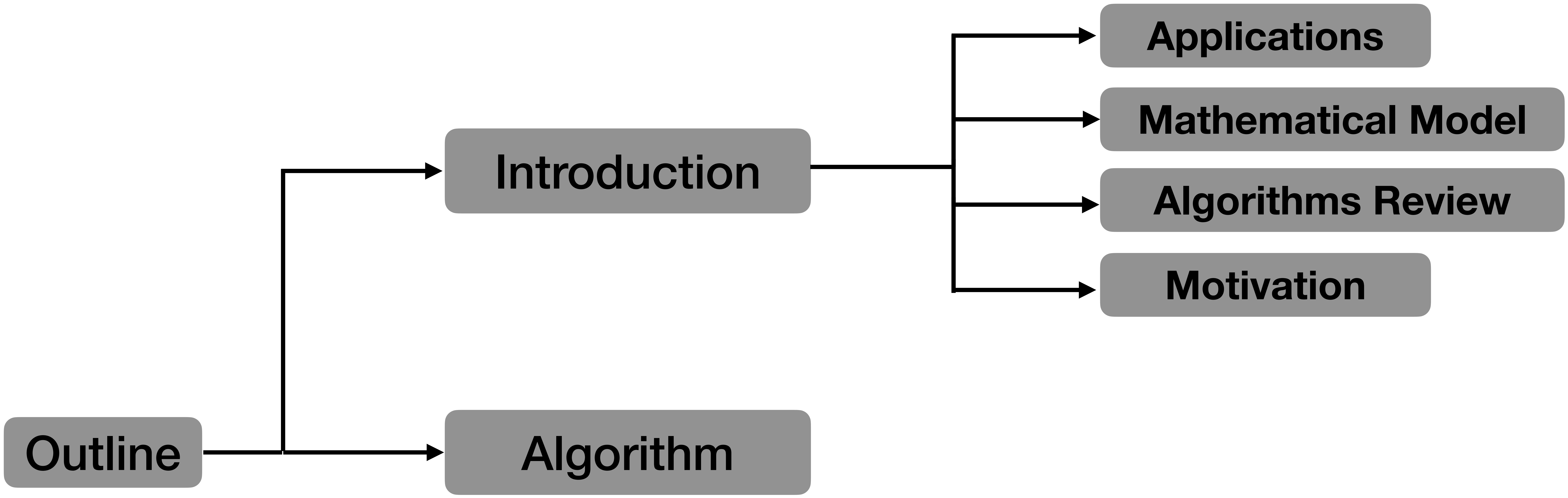
$\bar{x}_{k+1} = x_{k+1} + \alpha_{k+1}(x_{k+1} - x_k)$

until the stop criterion is satisfied.

Simple updating rule (without subproblem solving)

Lightweight parameter tuning burden

Improved convergence speed


Outline
Introduction
Algorithm
Applications
Mathematical Model
Algorithms Review
Motivation

$\mathbf{FP^2O}$

When $B \neq I$, the subproblem

$$\operatorname{Prox}_{\gamma g \circ B}(z) = \underset{u \in \mathbb{R}^d}{\arg\min}\{\gamma g \circ B(u) + \frac{1}{2}\|u - z\|_2^2\},$$

Enjoy the form of denoising problem

solved by $\mathbf{FP^2O}$

Algorithm: Fixed point algorithm based on proximity operator ($\mathbf{FP^2O}$)

Step 1 : Set $y_1 \in \mathbb{R}^r, 0 < \lambda < 1/\|B\|_2^2$

Step 2 : Perform the following fix point iteration and find fixed point y^*

$$y_{k+1} = (I - \operatorname{Prox}_{\frac{\gamma}{\lambda}g})(Bz + (I - \lambda BB^T)y_k) \longrightarrow \text{Fixed point iteration}$$

Step 3 : $\operatorname{Prox}_{\gamma g \circ B}(z) = z - \lambda By^*$

Micchelli, C. A., Shen, L., Xu, Y. (2011). Proximity algorithms for image models: denoising. *Inverse Problems*, *27*(4), 045009.

PDFP

$$\min_{x \in \mathbb{R}^d} \; f(x) + g \circ B(x),$$

Combining PGD with one step $\mathrm{FP^2O}$, one gets

Algorithm: Primal Dual Fixed Point

Step 1 : Set $x_1 \in \mathbb{R}^d$ **and choose proper** $0 < \gamma < 2/L_f$,

$0 \leq \lambda \leq 1/\|B\|_2^2$.

Step 2 : For $k = 1,2,\cdots$

$x_{k+\frac{1}{2}} = x_k - \gamma \nabla f(x_k)$ → One step GD

$x_{k+1} = \mathrm{Prox}_{\gamma g \circ B}(x_{k+\frac{1}{2}})$ → One step $\mathrm{FP^2O}$

until the stop criterion is satisfied.

Chen, P., Huang, J., Zhang, X. (2013). A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration. Inverse Problems, 29(2), 025011.

PDFP

$$\min_{x \in \mathbb{R}^d} \; f(x) + g \circ B(x),$$

Combining PGD with one step $\mathbf{FP^2O}$, one gets

Algorithm: Primal Dual Fixed Point method

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $0 < \gamma < 2/L_f$, $0 < \lambda < 1/\|B\|_2^2$.

Step 2 : For $k = 1,2,\cdots$

$$x_{k+\frac{1}{2}} = x_k - \gamma \nabla f(x_k)$$

One step GD

$$y_{k+1} = (I - \mathrm{Prox}_{\frac{\gamma}{\lambda} g})(Bx_{k+\frac{1}{2}} + (I - \lambda BB^T)y_k)$$

$$x_{k+1} = x_{k+\frac{1}{2}} - \lambda B^T y_{k+1}$$

One step $\mathrm{FP^2O}$

until the stop criterion is satisfied.

Chen, P., Huang, J., Zhang, X. (2013). A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration. Inverse Problems, 29(2), 025011.

PDFP

Algorithm: Primal Dual Fixed Point method

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $0 < \gamma < 2/L_f$, $0 < \lambda < 1/\|B\|_2^2$.

Step 2 : For $k = 1,2,\cdots$

$$x_{k+\frac{1}{2}} = x_k - \gamma \nabla f(x_k) \longrightarrow \text{One step GD}$$

$$\left.\begin{aligned} y_{k+1} &= (I - \mathrm{Prox}_{\frac{\gamma}{\lambda}g})(Bx_{k+\frac{1}{2}} + (I - \lambda BB^T)y_k) \\ x_{k+1} &= x_{k+\frac{1}{2}} - \lambda B^T y_{k+1} \end{aligned}\right\} \longrightarrow \text{One step FP}^2\text{O}$$

until the stop criterion is satisfied.

When $B = I, \lambda = 1$, PDFP becomes

$$x_{k+\frac{1}{2}} = x_k - \gamma \nabla f(x_k)$$

$$y_{k+1} = (I - \mathrm{Prox}_g)(x_{k+\frac{1}{2}})$$

$$x_{k+1} = x_k - \gamma \nabla f(x_k) - y_{k+1}$$

which is equivalent to

$$x_{k+1} = \mathrm{Prox}_{\gamma g}(x_k - \gamma \nabla f(x_k))$$

PDFP becomes PGD

PDFP

By using the change of variable $y_k := \frac{\lambda}{\gamma} y_k$ **and Moreau decomposition:**

$$z = \mathrm{Prox}_{cg}(z) + c\mathrm{Prox}_{\frac{g^*}{c}}(\frac{z}{c}),$$

one gets

Algorithm: PDFP (Loris-Verhoeven/PAPC)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $0 < \gamma < 2/L_f$
$0 < \lambda \leq 1/\|B\|_2^2$.
Step 2 : For $k = 1,2,\cdots$

$$\bar{x}_k = x_k - \gamma \nabla f(x_k) - \gamma B^T y_k$$

$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma} g^*}(\frac{\lambda}{\gamma} B\bar{x}_k + y_k)$$

$$x_{k+1} = x_k - \gamma \nabla f(x_k) - \gamma B^T y_{k+1}$$

until the stop criterion is satisfied.

Loris, I., Verhoeven, C. (2011). On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty. Inverse Problems, 27(12), 125007.
Drori, Y., Sabach, S., & Teboulle, M. (2015). A simple algorithm for a class of nonsmooth convex–concave saddle-point problems. Operations Research Letters, 43(2), 209-214.

PDFP

Algorithm: PDFP (Loris-Verhoeven/PAPC)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $0 < \gamma < 2/L_f$ $0 < \lambda \leq 1/\|B\|_2^2$.

Step 2 : For $k = 1,2,\cdots$

$$\boxed{\bar{x}_k = x_k - \gamma \nabla f(x_k) - \gamma B^T y_k}$$

$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma} g^*}(-\frac{\lambda}{\gamma} B\bar{x}_k + y_k)$$

$$\boxed{x_{k+1} = x_k - \gamma \nabla f(x_k) - \gamma B^T y_{k+1}}$$

until the stop criterion is satisfied.

$$\mathcal{G}(\bar{x}_k, \bar{y}_k) \leq \mathcal{O}\left(\boxed{\frac{L_f}{k}}\right) + \mathcal{O}\left(\frac{\rho_{\max}(I - \lambda BB^T)}{k}\right)$$

Ya-Nan Zhu, Xiaoqun Zhang. 2021. A stochastic variance reduced primal dual fixed point method for linearly constrained separable optimization. SIAM Journal on Imaging Sciences, 14(3), 1326-1353.

PDFP + Acceleration

IPDFP

Algorithm: Inertial Primal dual fixed point method (IPDFP)

Step 1 : Set $x_1 \in \mathbb{R}^d, p_1 \in \mathbb{R}^r$ **and choose proper** $0 < \gamma < 2/L_f$, $0 < \lambda < 1/\|B\|^2$ **and let** $M = I - \lambda BB^T$

Step 2 : For $k = 1,2,\cdots$

$$z_k = x_k + \alpha_k(x_k - x_{k-1})$$
$$v_k = y_k + \alpha_k M(y_k - y_{k-1})$$

$\rightarrow$ **inertial**

$$\bar{x}_{k+1} = z_k - \gamma \nabla f(z_k) - \gamma B^T y_k$$
$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma} g^*}\left(\frac{\lambda}{\gamma} B\bar{x}_{k+1} + v_k\right)$$
$$x_{k+1} = z_k - \gamma \nabla f(z_k) - \gamma B^T y_{k+1}$$

$\rightarrow$ **PDFP**

until the stop criterion is satisfied.

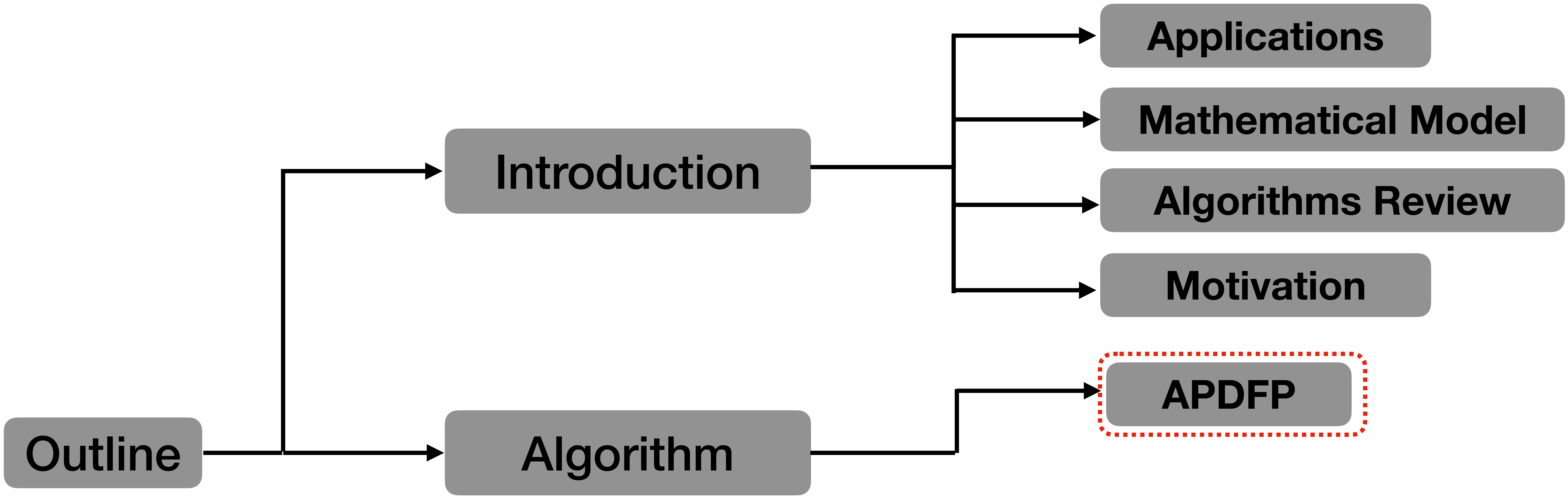
Ya-Nan Zhu and Xiaoqun Zhang. "Two Modified Schemes for the Primal Dual Fixed Point Method". CSIAM Trans. Appl. Math., 2 (2021), pp. 108-130.

IPDFP

Algorithm: Inertial Primal dual fixed point method (IPDFP)
Step 1 : Set $x_1 \in \mathbb{R}^d, p_1 \in \mathbb{R}^r$ **and choose proper** $0 < \gamma < 2/L_f$, $0 < \lambda < 1/\|B\|^2$ **and let** $M = I - \lambda BB^T$ **Step 2 : For** $k = 1,2,\cdots$ $z_k = x_k + \alpha_k(x_k - x_{k-1})$ $v_k = y_k + \alpha_k M(y_k - y_{k-1})$ → **inertial** $\bar{x}_{k+1} = z_k - \gamma \nabla f(z_k) - \gamma B^T y_k$ $y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma} g^*}(\frac{\lambda}{\gamma} B\bar{x}_{k+1} + v_k)$ → **PDFP** $x_{k+1} = z_k - \gamma \nabla f(z_k) - \gamma B^T y_{k+1}$ **until the stop criterion is satisfied.**

The optimal inertial parameter α_k is problem dependent

Only convergence, no convergence rate


Outline
Introduction
Algorithm
Applications
Mathematical Model
Algorithms Review
Motivation
APDFP

APDFP

Algorithm: Accelerated Primal Dual Fixed Point method (APDFP)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** γ_k, λ**.**
Step 2 : For $k = 1,2,\cdots$

$$\theta_k = 2/(k+1)$$

$$\tilde{x}_k = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_k$$

$$\bar{x}_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_k$$

$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma_k}\theta_k g^*}\left(\frac{\lambda}{\gamma_k}\theta_k B \bar{x}_{k+1} + y_k\right)$$

$$x_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_{k+1}$$

$$x_{k+1}^{\mathrm{ag}} = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_{k+1}$$

until the stop criterion is satisfied.

PDFP&APDFP

Algorithm: PDFP (Loris-Verhoeven/PAPC)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $0 < \gamma < 2/L_f$
$0 < \lambda \leq 1/\|B\|_2^2$.
Step 2 : For $k = 1,2,\cdots$

$$\bar{x}_k = x_k - \gamma \nabla f(x_k) - \gamma B^T y_k$$
$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma} g^*}\left(\frac{\lambda}{\gamma} B \bar{x}_k + y_k\right)$$
$$x_{k+1} = x_k - \gamma \nabla f(x_k) - \gamma B^T y_{k+1}$$

until the stop criterion is satisfied.

Algorithm: Accelerated Primal Dual Fixed Point method (APDFP)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** γ_k, λ.
Step 2 : For $k = 1,2,\cdots$

$$\theta_k = 2/(k+1)$$
$$\tilde{x}_k = (1-\theta_k) x_k^{\mathrm{ag}} + \theta_k x_k$$
$$\bar{x}_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_k$$
$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma_k}\theta_k g^*}\left(\frac{\lambda}{\gamma_k}\theta_k B \bar{x}_{k+1} + y_k\right)$$
$$x_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_{k+1}$$
$$x_{k+1}^{\mathrm{ag}} = (1-\theta_k) x_k^{\mathrm{ag}} + \theta_k x_{k+1}$$

until the stop criterion is satisfied.

PDFP&APDFP

Algorithm: PDFP (Loris-Verhoeven/PAPC)

Step 1 : **Set** $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $0 < \gamma < 2/L_f$ $0 < \lambda \leq 1/\|B\|_2^2$.
Step 2 : For $k = 1,2,\cdots$

$$\bar{x}_k = x_k - \gamma \nabla f(x_k) - \gamma B^T y_k$$

$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma} g^*}(\frac{\lambda}{\gamma} B\bar{x}_k + y_k)$$

$$x_{k+1} = x_k - \gamma \nabla f(x_k) - \gamma B^T y_{k+1}$$

until the stop criterion is satisfied.

Algorithm: Accelerated Primal Dual Fixed Point method (APDFP)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** γ_k, λ.
Step 2 : For $k = 1,2,\cdots$

$$\theta_k = 2/(k+1)$$

$$\tilde{x}_k = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_k$$

$$\bar{x}_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_k$$

$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma_k}\theta_k g^*}(\frac{\lambda}{\gamma_k}\theta_k B\bar{x}_{k+1} + y_k)$$

$$x_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_{k+1}$$

$$x_{k+1}^{\mathrm{ag}} = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_{k+1}$$

until the stop criterion is satisfied.

PDFP&APDFP

Algorithm: PDFP (Loris-Verhoeven/PAPC)
Step 1 : **Set** $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $0 < \gamma < 2/L_f$ $0 < \lambda \leq 1/\|B\|_2^2$.
Step 2 : **For** $k = 1,2,\cdots$
$\bar{x}_k = x_k - \gamma \nabla f(x_k) - \gamma B^T y_k$
$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma} g^*}(\frac{\lambda}{\gamma} B\bar{x}_k + y_k)$
$x_{k+1} = x_k - \gamma \nabla f(x_k) - \gamma B^T y_{k+1}$
until the stop criterion is satisfied.

Algorithm: Accelerated Primal Dual Fixed Point method (APDFP)
Step 1 : **Set** $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** γ_k, λ.
Step 2 : **For** $k = 1,2,\cdots$
$\theta_k = 2/(k+1)$
$\tilde{x}_k = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_k$
$\bar{x}_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_k$
$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma_k}\theta_k g^*}(\frac{\lambda}{\gamma_k}\theta_k B\bar{x}_{k+1} + y_k)$
$x_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_{k+1}$
$x_{k+1}^{\mathrm{ag}} = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_{k+1}$
until the stop criterion is satisfied.

PDFP&APDFP

Algorithm: PDFP (Loris-Verhoeven/PAPC)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $0 < \gamma < 2/L_f$ $0 < \lambda \le 1/\|B\|_2^2$.
Step 2 : For $k = 1,2,\cdots$

$$\bar{x}_k = x_k - \gamma \nabla f(\boxed{x_k}) - \gamma B^T y_k$$

$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma} g^*}\left(\frac{\lambda}{\gamma} B\bar{x}_k + y_k\right)$$

$$x_{k+1} = x_k - \gamma \nabla f(\boxed{x_k}) - \gamma B^T y_{k+1}$$

until the stop criterion is satisfied.

Algorithm: Accelerated Primal Dual Fixed Point method (APDFP)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** γ_k, λ.
Step 2 : For $k = 1,2,\cdots$

$$\theta_k = 2/(k+1)$$

$$\tilde{x}_k = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_k$$

$$\bar{x}_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_k$$

$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma_k}\theta_k g^*}\left(\frac{\lambda}{\gamma_k}\theta_k B\bar{x}_{k+1} + y_k\right)$$

$$x_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_{k+1}$$

$$x_{k+1}^{\mathrm{ag}} = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_{k+1}$$

until the stop criterion is satisfied.

PDFP&APDFP

Algorithm: PDFP (Loris-Verhoeven/PAPC)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** $0 < \gamma < 2/L_f$ $0 < \lambda \leq 1/\|B\|_2^2$.
Step 2 : For $k = 1,2,\cdots$

$$\bar{x}_k = x_k - \gamma \nabla f(x_k) - \gamma B^T y_k$$

$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma}g^*}(\frac{\lambda}{\gamma}B\bar{x}_k + y_k)$$

$$x_{k+1} = x_k - \gamma \nabla f(x_k) - \gamma B^T y_{k+1}$$

until the stop criterion is satisfied.

Algorithm: Accelerated Primal Dual Fixed Point method (APDFP)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ **and choose proper** γ_k, λ.
Step 2 : For $k = 1,2,\cdots$

$$\theta_k = 2/(k+1)$$

$$\tilde{x}_k = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_k$$

$$\bar{x}_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_k$$

$$y_{k+1} = \mathrm{Prox}_{\frac{\lambda}{\gamma_k}\theta_k g^*}(\frac{\lambda}{\gamma_k}\theta_k B\bar{x}_{k+1} + y_k)$$

$$x_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_{k+1}$$

$$x_{k+1}^{\mathrm{ag}} = (1-\theta_k)x_k^{\mathrm{ag}} + \theta_k x_{k+1}$$

until the stop criterion is satisfied.

APDFP $\xrightarrow{B=I,\lambda=1}$ NAG

APDFP

$$\begin{aligned}
\theta_k &= 2/(k+1)\\
\tilde{x}_k &= (1-\theta_k)x_k^{\text{ag}} + \theta_k x_k\\
\bar{x}_{k+1} &= x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_k\\
y_{k+1} &= \text{Prox}_{\frac{\lambda}{\gamma_k}\theta_k g^*}\Big(\frac{\lambda}{\gamma_k}\theta_k B \bar{x}_{k+1} + y_k\Big)\\
x_{k+1} &= x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_{k+1}\\
x_{k+1}^{\text{ag}} &= (1-\theta_k)x_k^{\text{ag}} + \theta_k x_{k+1}
\end{aligned}$$

$B = I, \lambda = 1$

$$\begin{aligned}
\theta_k &= 2/(k+1)\\
\tilde{x}_k &= (1-\theta_k)x_k^{\text{ag}} + \theta_k x_k\\
\bar{x}_{k+1} &= x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k y_k\\
y_{k+1} &= \text{Prox}_{\frac{\theta_k}{\gamma_k} g^*}\Big(\frac{\theta_k}{\gamma_k} \bar{x}_{k+1} + y_k\Big)\\
x_{k+1} &= x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k y_{k+1}\\
x_{k+1}^{\text{ag}} &= (1-\theta_k)x_k^{\text{ag}} + \theta_k x_{k+1}
\end{aligned}$$

Simplify

$$\begin{aligned}
\theta_k &= 2/(k+1)\\
\tilde{x}_k &= (1-\theta_k)x_k^{\text{ag}} + \theta_k x_k\\
y_{k+1} &= \text{Prox}_{\frac{\theta_k}{\gamma_k} g^*}\Big(\frac{\theta_k}{\gamma_k}(x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k))\Big)\\
x_{k+1} &= x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k y_{k+1}\\
x_{k+1}^{\text{ag}} &= (1-\theta_k)x_k^{\text{ag}} + \theta_k x_{k+1}
\end{aligned}$$

Simplify

$$\begin{aligned}
\theta_k &= 2/(k+1)\\
\tilde{x}_k &= (1-\theta_k)x_k^{\text{ag}} + \theta_k x_k\\
x_{k+1} &= x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k \text{Prox}_{\frac{\theta_k}{\gamma_k} g^*}\Big(\frac{\theta_k}{\gamma_k}(x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k))\Big)\\
x_{k+1}^{\text{ag}} &= (1-\theta_k)x_k^{\text{ag}} + \theta_k x_{k+1}
\end{aligned}$$

Moreau Decomposition

$$q = \text{Prox}_{\frac{\gamma_k}{\theta_k} g}(q) + \gamma_k/\theta_k \text{Prox}_{\frac{\theta_k}{\gamma_k} g^*}\Big(\frac{\theta_k}{\gamma_k} q\Big),$$

NAG

$$\begin{aligned}
\theta_k &= 2/(k+1)\\
\tilde{x}_k &= (1-\theta_k)x_k^{\text{ag}} + \theta_k x_k\\
x_{k+1} &= \text{Prox}_{(\gamma/\theta_k)g}(x_k - \gamma/\theta_k \nabla f(\tilde{x}_k))\\
x_{k+1}^{\text{ag}} &= (1-\theta_k)x_k^{\text{ag}} + \theta_k x_{k+1}
\end{aligned}$$


Outline
Introduction
Applications
Mathematical Model
Algorithms Review
Motivation
Algorithm
APDFP
Convergence Results

Convergence

Definition

Let $\tilde{z} = (\tilde{x}, \tilde{y})$ and $z = (x, y)$, define

$$Q(\tilde{z}, z) = [f(\tilde{x}) + \langle B\tilde{x}, y\rangle - g^*(y)] - [f(x) + \langle Bx, \tilde{y}\rangle - g^*(\tilde{y})] \, .$$

Define the partial primal dual gap as

$$\begin{aligned}\mathcal{G}_{B_1 \times B_2}(\tilde{x}, \tilde{y}) &= max_{z \in B_1 \times B_2} Q(\tilde{z}, z) \\ &= max_{y \in B_2}[f(\tilde{x}) + \langle B\tilde{x}, y\rangle - g^*(y)] - min_{x \in B_1}[f(x) + \langle Bx, \tilde{y}\rangle - g^*(\tilde{y})]\end{aligned}$$

where the $B_1 \times B_2$ are bounded set containing the saddle point of minmax formulation

Convergence

$$\min_{x\in\mathbb{R}^d} \; f(x) + g \circ B(x),$$

Theorem

Suppose the function $f(x)$ is L_f smooth convex function and g is convex Lipchitz continuous. $x_{k+1}^{\text{ag}}, y_{k+1}^{\text{ag}}$ be the iterate in APDFP, choose parameters such that

(1) $0 < \lambda \leq 1/\rho_{\max}(BB^T)$

(2) $\gamma_k = \dfrac{1}{L_f + ck}, 0 < c \leq L_f$

Then the following inequality holds

$$\mathcal{G}_{B_1\times B_2}(x_{k+1}^{\text{ag}}, y_{k+1}^{\text{ag}}) \leq \frac{2L_f}{(k+1)^2}\Omega_1 + \frac{2ck}{(k+1)^2}\Omega_1 + \frac{1}{2(L_f+ck)}\frac{\rho_{\max}(I-\lambda BB^T)}{\lambda}\Omega_2 .$$

where Ω_1, Ω_2 are constants related to B_1, B_2, respectively.

Convergence

$$\min_{x\in\mathbb{R}^d} \; f(x) + g \circ B(x),$$

Theorem

Suppose the function $f(x)$ is L_f smooth convex function and g is convex Lipchitz continuous. $x_{k+1}^{\text{ag}}, y_{k+1}^{\text{ag}}$ be the iterate in APDFP, choose parameters such that

(1) $0 < \lambda \le 1/\rho_{\max}(BB^T)$

(2) $\gamma_k = \dfrac{1}{L_f + ck}, 0 < c \le L_f$

Then the following inequality holds

$$\mathcal{G}_{B_1\times B_2}(x_{k+1}^{\text{ag}}, y_{k+1}^{\text{ag}}) \le \boxed{\frac{2L_f}{(k+1)^2}}\Omega_1 + \frac{2ck}{(k+1)^2}\Omega_1 + \frac{1}{2(L_f+ck)}\frac{\rho_{\max}(I-\lambda BB^T)}{\lambda}\Omega_2 .$$

where Ω_1, Ω_2 are constants related to B_1, B_2, respectively.

Convergence

$$\min_{x\in\mathbb{R}^d} \ f(x) + g \circ B(x),$$

Theorem

Suppose the function $f(x)$ is L_f smooth convex function and g is convex Lipchitz continuous. $x_{k+1}^{\text{ag}}, y_{k+1}^{\text{ag}}$ be the iterate in APDFP, choose parameters such that

(1) $0 < \lambda \leq 1/\rho_{\max}(BB^T)$

(2) $\gamma_k = \dfrac{1}{L_f + ck}, 0 < c \leq L_f$

Then the following inequality holds

$$\mathcal{G}_{B_1\times B_2}(x_{k+1}^{\text{ag}}, y_{k+1}^{\text{ag}}) \leq \frac{2L_f}{(k+1)^2}\Omega_1 + \frac{2ck}{(k+1)^2}\Omega_1 + \frac{1}{2(L_f+ck)}\frac{\rho_{\max}(I-\lambda BB^T)}{\lambda}\Omega_2 .$$

where Ω_1, Ω_2 are constants related to B_1, B_2, respectively.

Convergence

$$\min_{x\in\mathbb{R}^d} \; f(x) + g \circ B(x),$$

Theorem

Suppose the function $f(x)$ is L_f smooth convex function and g is convex Lipchitz continuous. $x_{k+1}^{\text{ag}}, y_{k+1}^{\text{ag}}$ be the iterate in APDFP, choose parameters such that

(1) $0 < \lambda \leq 1/\rho_{\max}(BB^T)$

(2) $\gamma_k = \dfrac{1}{L_f + ck}, 0 < c \leq L_f$

Then the following inequality holds

$$\mathcal{G}_{B_1\times B_2}(x_{k+1}^{\text{ag}}, y_{k+1}^{\text{ag}}) \leq \frac{2L_f}{(k+1)^2}\Omega_1 + \frac{2ck}{(k+1)^2}\Omega_1 + \frac{1}{2(L_f+ck)}\frac{\rho_{\max}(I-\lambda BB^T)}{\lambda}\Omega_2 \,.$$

where Ω_1, Ω_2 are constants related to B_1, B_2, respectively.

Empirically setting $c = 0$ does not ruin the convergence (speed)

LP-ADMM

Algorithm: Linearized Preconditioned ADMM (LP-ADMM)

Step 1 : Set $x_1, z_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ and choose proper $C > 0$, let $\rho_k := \rho = C/\|B\|, \gamma_k := \gamma = 1/(L_f + \rho\|B\|^2)$.

Step 2 : For $k = 1,2,\cdots$

$$x_{k+1} = x_k - \gamma_k(\nabla f(x_k) + \rho_k B^T(Bx_k - z_k) + B^T y_k)$$

$$z_{k+1} = \text{Prox}_{g/\rho_k}(Bx_{k+1} + y_k/\rho_k)$$

$$y_{k+1} = y_k + \rho_k(Bx_{k+1} - z_{k+1})$$

until the stop criterion is satisfied.

Nesterov

Algorithm: Accelerated ADMM (AADMM)

Step 1 : Set $x_1, z_1, \tilde{x}_1, \bar{x}_1, \in \mathbb{R}^d, p_1 \in \mathbb{R}^m$ and choose proper $\rho > 0, \sigma_k = (k-1)\rho/k, \gamma_k = k/(2/L_f + \rho k\|B\|^2)$

Step 2 : For $k = 1,2,\cdots$

$$\theta_k = 2/(k+1)$$

$$\tilde{x}_k = (1-\theta_k)x_k^{ag} + \theta_k x_k$$

$$x_{k+1} = x_k - \gamma_k(\nabla f(\tilde{x}_k) + \sigma_k B^T(Bx_k - z_k) + B^T y_k)$$

$$x_{k+1}^{ag} = (1-\theta_k)x_k^{ag} + \theta_k z_{k+1}$$

$$z_{k+1} = \text{Prox}_{g/\rho}(Bx_{k+1} + y_k/\rho)$$

$$y_{k+1} = y_k + \rho(Bx_{k+1} - z_{k+1})$$

until the stop criterion is satisfied.

LPDHGm

Algorithm: Linearized Primal Dual Hybrid Gradient (LPDHGm)

Step 1 : Set $x_1, \bar{x}_1, \in \mathbb{R}^d, y_1 \in \mathbb{R}^m$ and choose proper $0 < \alpha_k = \sigma_{k-1}/\sigma_k = \gamma_{k-1}/\gamma_k \le 1,\ L_f\sigma_k + \|B\|^2\gamma_k\sigma_k \le 1,$.

Step 2 : For $k = 1,2,\cdots$

$$y_{k+1} = \text{Prox}_{\sigma_k g^*}(y_k + \sigma_k B\bar{x}_k)$$

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k) - \gamma_k B^T y_{k+1}$$

$$\bar{x}_{k+1} = x_{k+1} + \alpha_k(x_{k+1} - x_k)$$

until the stop criterion is satisfied.

Nesterov

Algorithm: APD

Step 1 : Set $x_1, z_1, \tilde{x}_1, \bar{x}_1, \in \mathbb{R}^d, y_1 \in \mathbb{R}^m$ and choose proper $C > 0, \tau_k = C/\|B\|, \gamma_k = k/(2L_f + k\|B\|C)$

Step 2 : For $k = 1,2,\cdots$

$$\theta_k = 2/(k+1), \alpha_k = (k-1)/k$$

$$y_{k+1} = \text{Prox}_{\tau_k g^*}(y_k + \tau_k B\bar{x}_k)$$

$$\tilde{x}_k = (1-\theta_k)x_k^{ag} + \theta_k x_k$$

$$x_{k+1} = x_k - \gamma_k \nabla f(\tilde{x}_k) - \gamma_k B^T y_{k+1}$$

$$x_{k+1}^{ag} = (1-\theta_k)x_k^{ag} + \theta_k x_{k+1}$$

$$\bar{x}_{k+1} = x_{k+1} + \alpha_k(x_{k+1} - x_k)$$

until the stop criterion is satisfied.

PDFP

Algorithm: PDFP (Loris-Verhoeven)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ and choose proper $0 < \gamma < 2/L_f$ $0 \le \lambda \le 1/\|B\|_2^2$.

Step 2 : For $k = 1,2,\cdots$

$$\bar{x}_k = x_k - \gamma\nabla f(x_k) - \gamma B^T y_k$$

$$y_{k+1} = \text{Prox}_{\frac{\lambda}{\gamma}g^*}(\frac{\lambda}{\gamma}B\bar{x}_k + y_k)$$

$$x_{k+1} = x_k - \gamma\nabla f(x_k) - \gamma B^T y_{k+1}$$

until the stop criterion is satisfied.

Nesterov

Algorithm: Accelerated Primal Dual Fixed Point method (APDFP)

Step 1 : Set $x_1 \in \mathbb{R}^d, y_1 \in \mathbb{R}^r$ and choose proper γ_k, λ.

Step 2 : For $k = 1,2,\cdots$

$$\theta_k = 2/(k+1)$$

$$\tilde{x}_k = (1-\theta_k)x_k^{ag} + \theta_k x_k$$

$$\bar{x}_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_k$$

$$y_{k+1} = \text{Prox}_{\frac{\lambda}{\gamma_k}\theta_k g^*}(\frac{\lambda}{\gamma_k}\theta_k B\bar{x}_{k+1} + y_k)$$

$$x_{k+1} = x_k - \gamma_k/\theta_k \nabla f(\tilde{x}_k) - \gamma_k/\theta_k B^T y_{k+1}$$

$$x_{k+1}^{ag} = (1-\theta_k)x_k^{ag} + \theta_k x_{k+1}$$

until the stop criterion is satisfied.


Outline
Introduction
Applications
Mathematical Model
Algorithms Review
Motivation
Algorithm
APDFP
Convergence Results
Numerical Experiments

Graph-Guided Logistic Regression

The optimization problem is given as follows:

$$\min_{x\in\mathbb{R}^d} \frac{1}{n}\sum_{i=1}^{n} f_i(x) + \mu_1\|x\|_2^2 + \mu_2\|Bx\|_1,$$

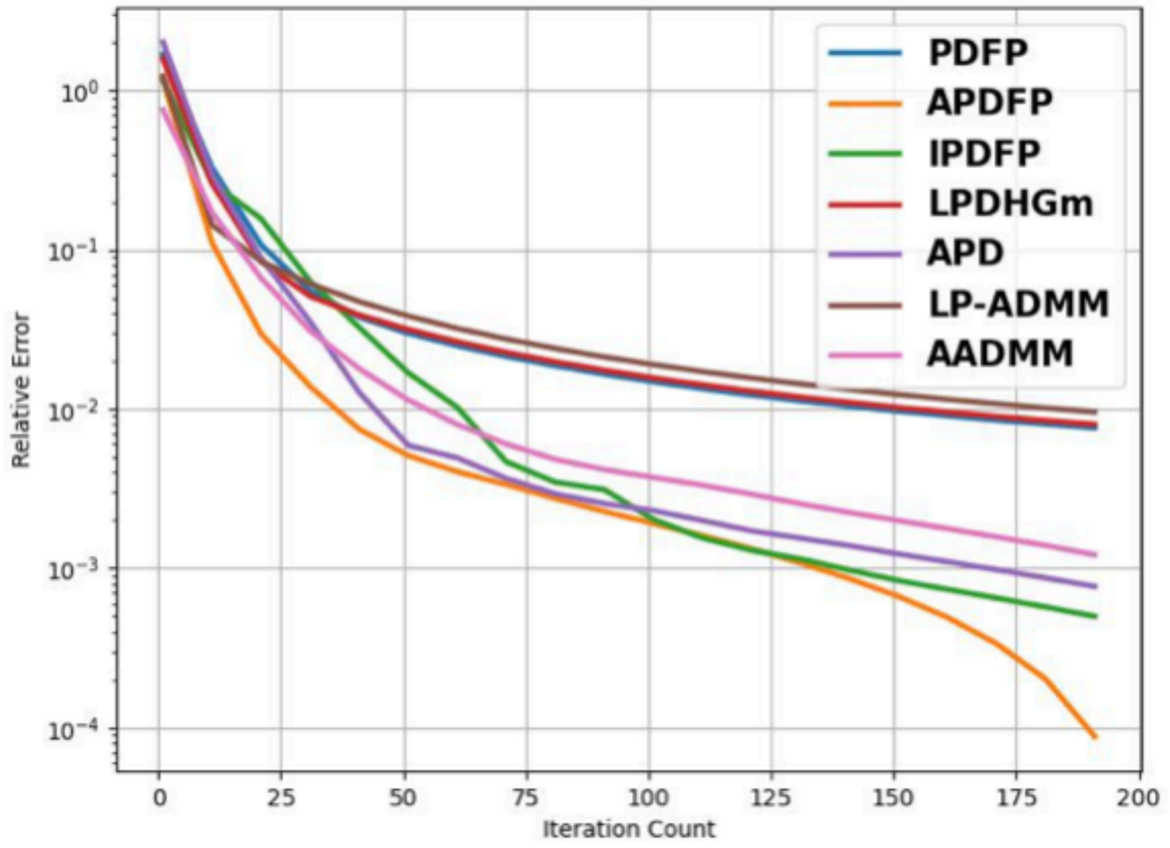
where

- $f_i(x) = \log(1 + \exp(-b_i a_i^T x), i = 1,\cdots,n$ **is the hinge loss and b_i denote the label of the i-th sample a_i.**
- **The matrix B is determined by sparse inverse covariance selection**
- μ_1, μ_2 **are regularization parameters.**
- **Dataset: *a9a, Mushroom, w8a* (https://www.csie.ntu.edu.tw/~cjlin/libsvm/).**

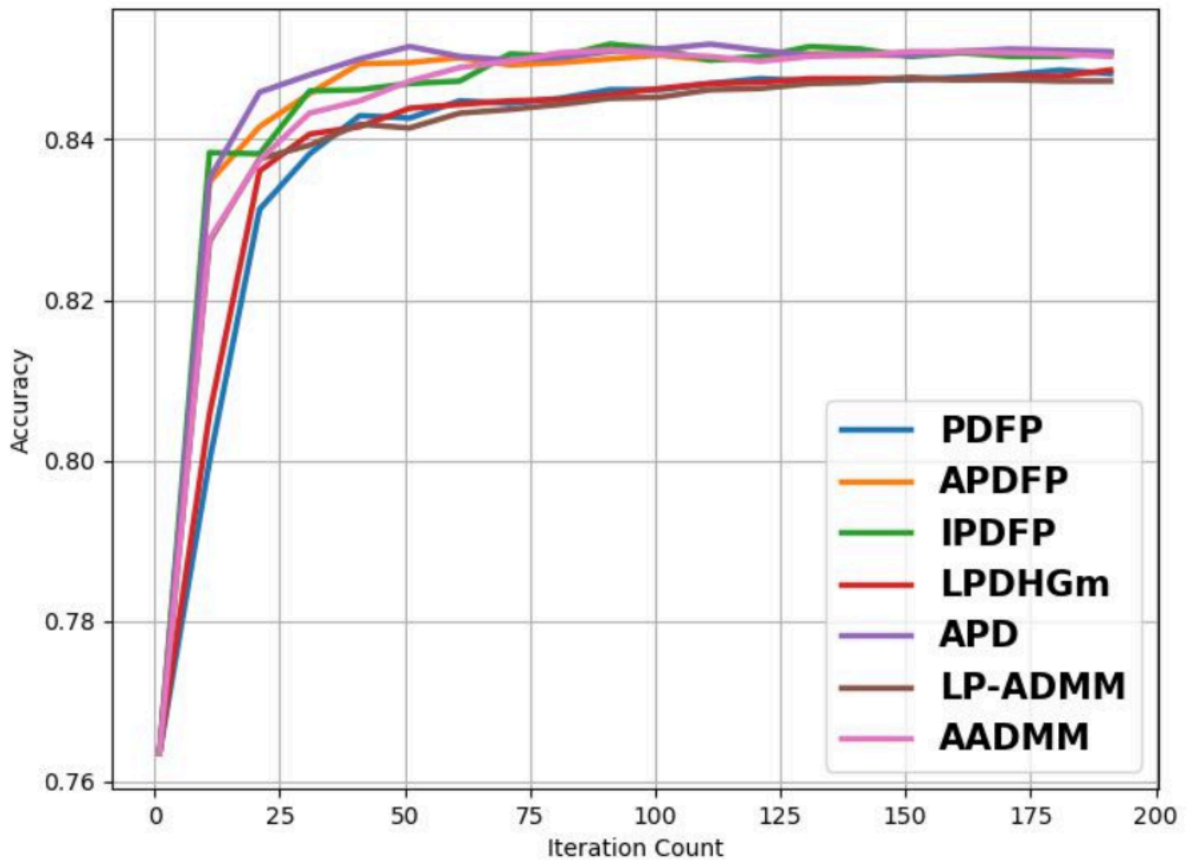
Jerome Friedman, Trevor Hastie, and Robert Tibshirani. “Sparse inverse covariance estimation with the graphical lasso”. In: Biostatistics 9.3 (2008), pp. 432–441.

Graph-Guided Logistic Regression

a9a

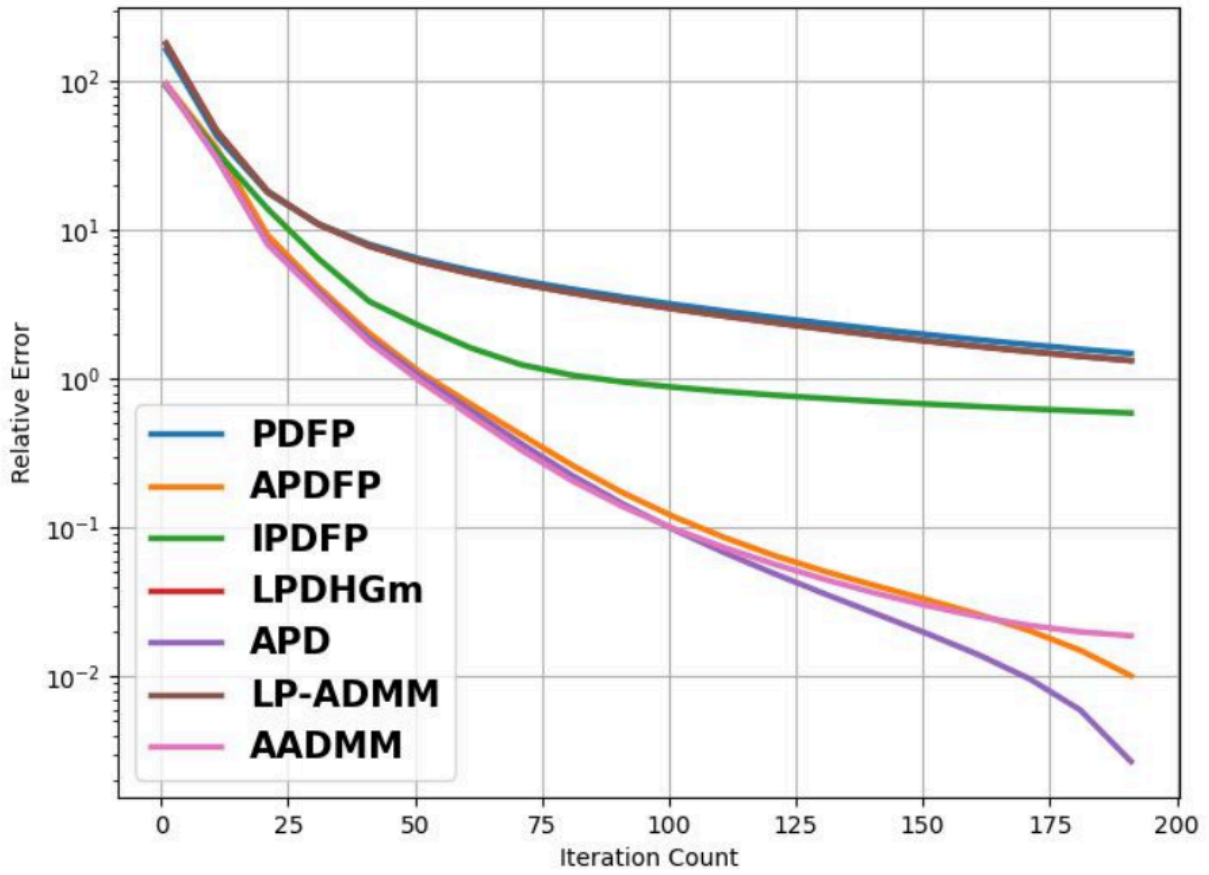


Relative error of Objective v.s. iteration

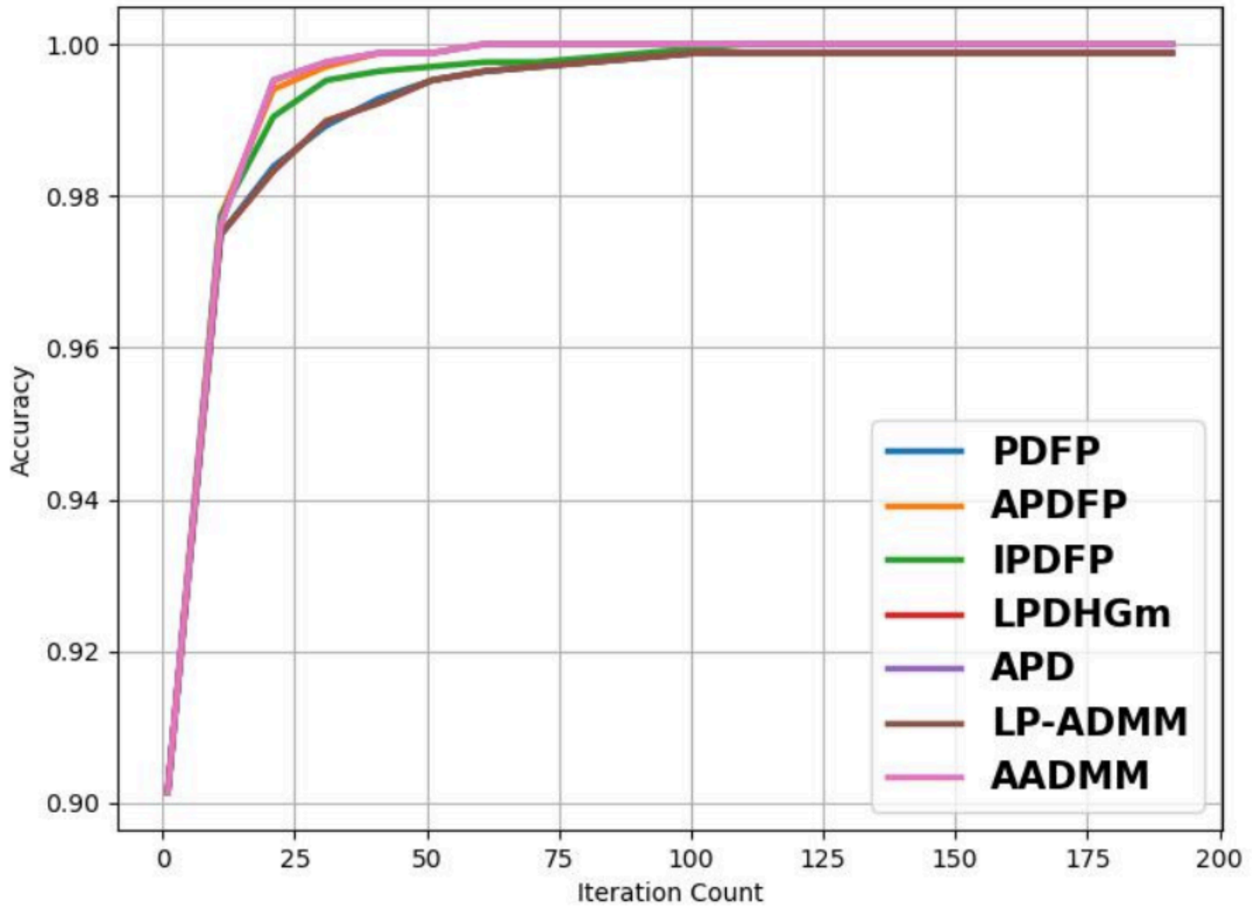


Testing Accuracy v.s. iteration

mushroom

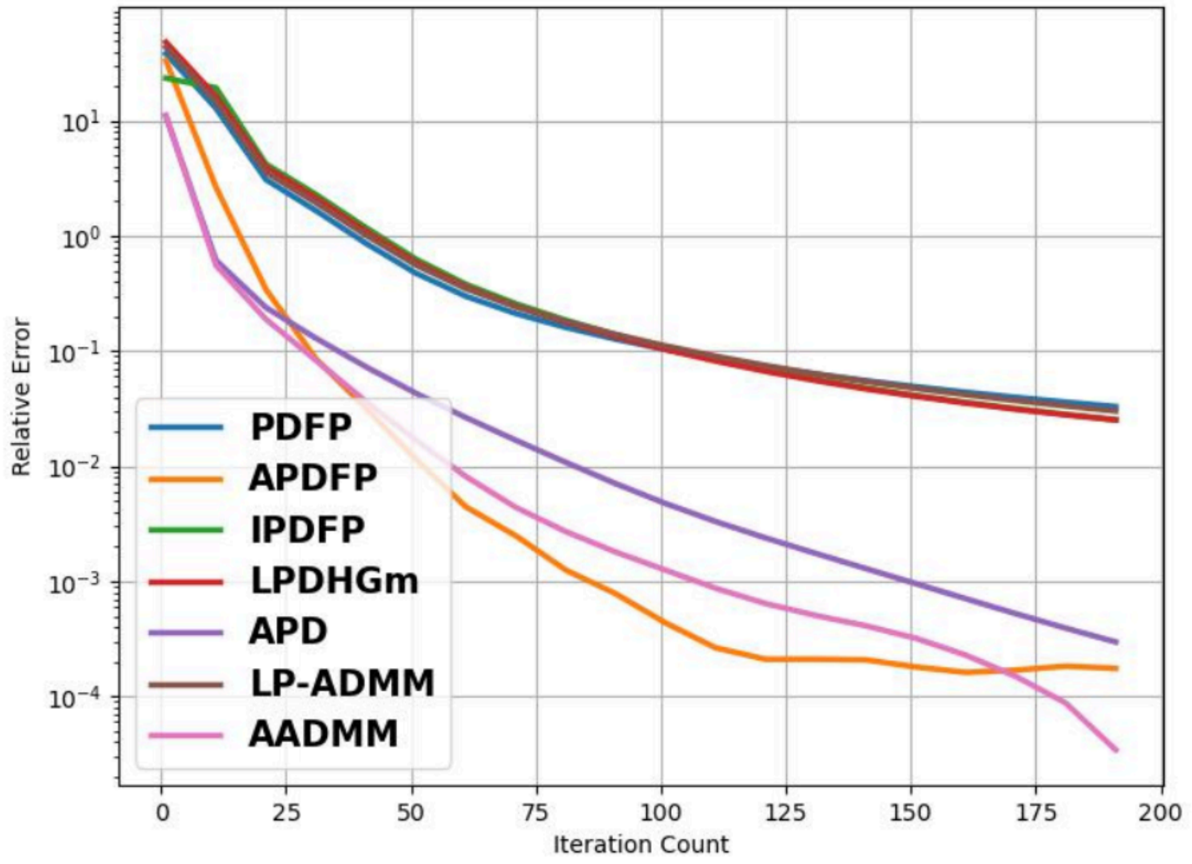


Relative error of Objective v.s. iteration

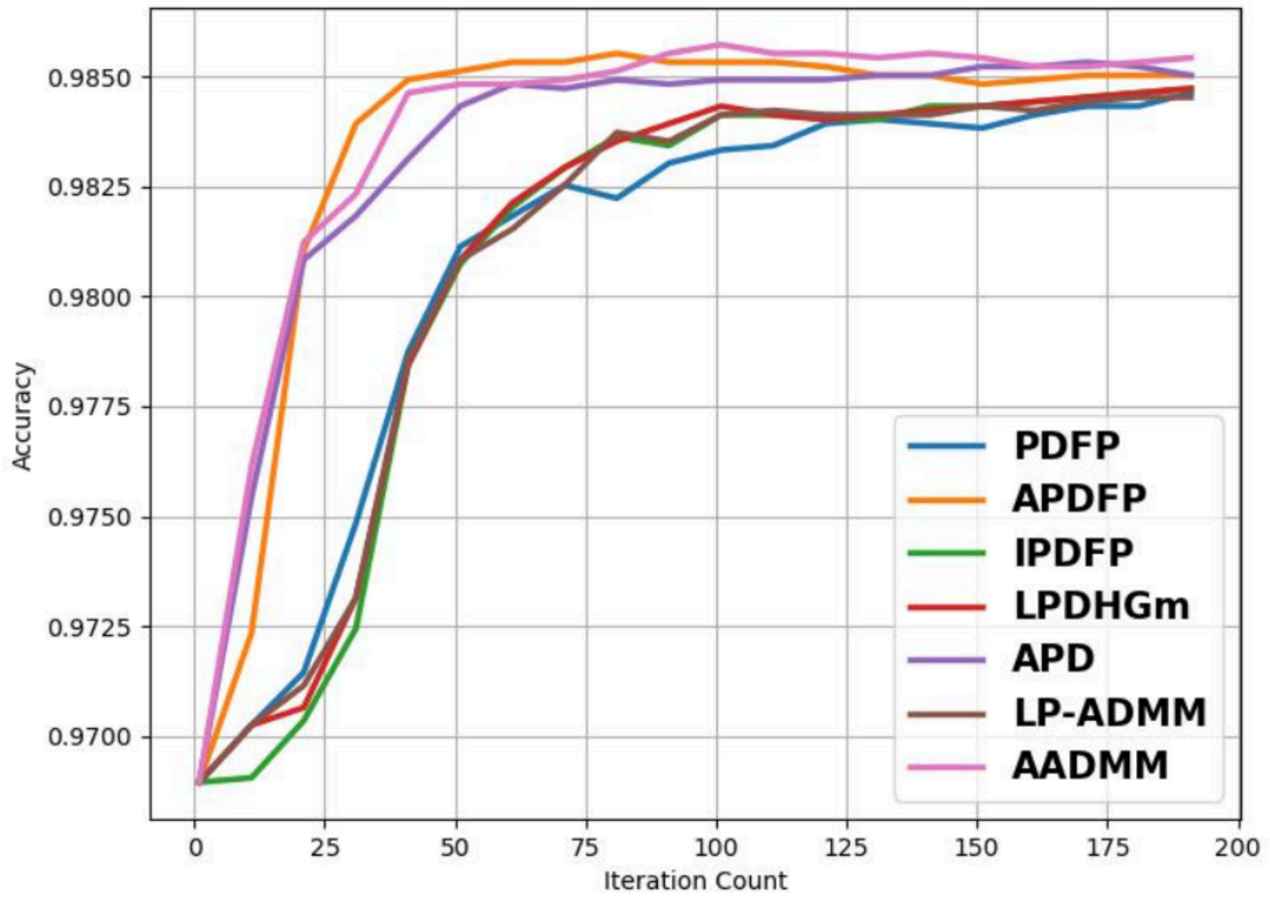


Testing Accuracy v.s. iteration

W8a



Relative error of Objective v.s. iteration



Testing Accuracy v.s. iteration

2D computerized tomography (CT) reconstruction

We consider the CT reconstruction using TV-L_2 model i.e.

$$\arg\min_{x\in\mathbb{R}^d} \frac{1}{2}\|\mathscr{A}x - b\|_2^2 + \mu\|\nabla x\|_{2,1},$$

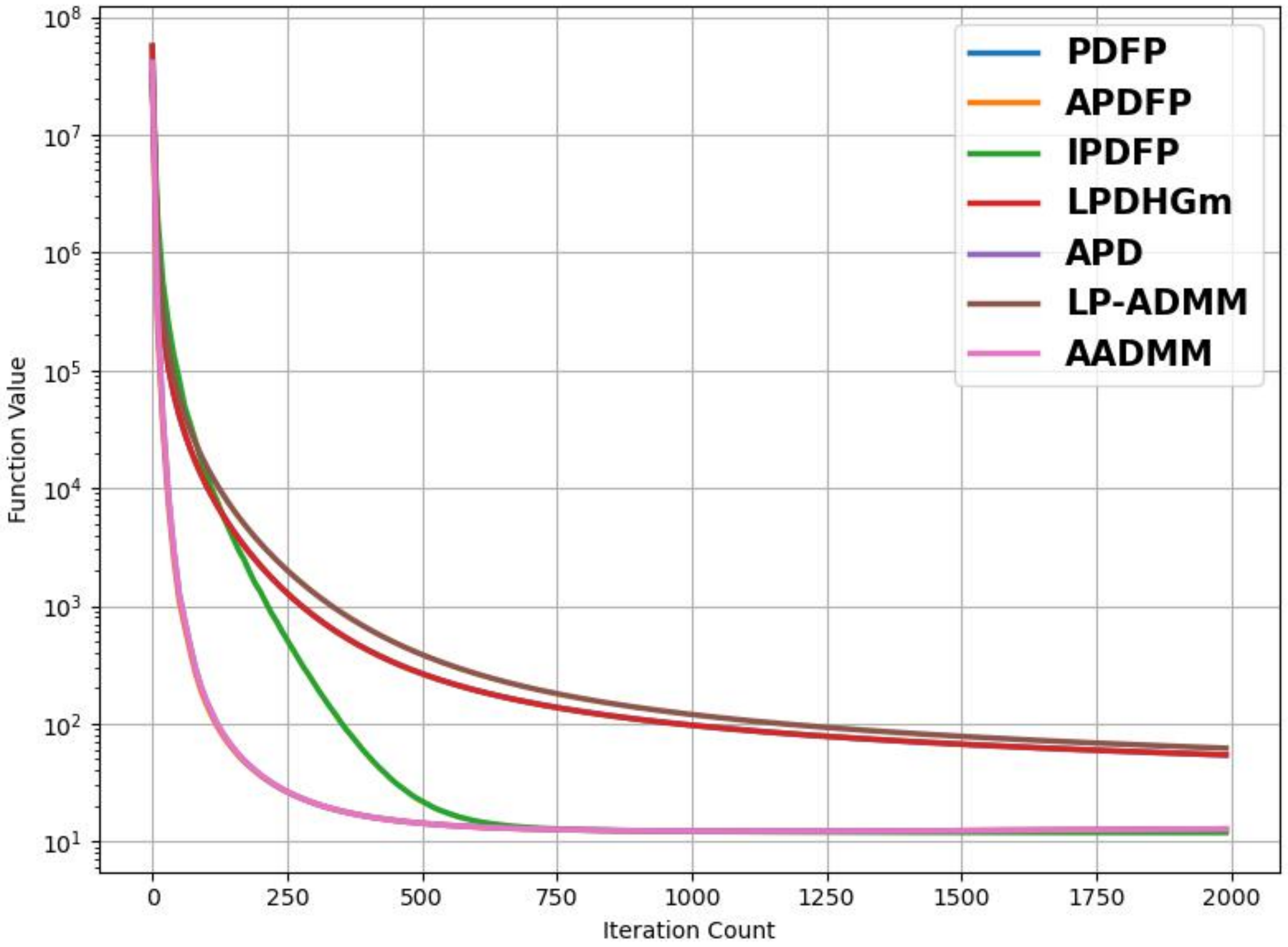
where

- $\mathscr{A}$ **is discrete X-ray* transform using fan beam geometry (Detectors line: 512, number of viewers: 360)**
- $b \in \mathbb{R}^{512\times 360}$ **: Measured projections**
- $x \in \mathbb{R}^{512\times 512}$ **: Image to be reconstructed**
- ∇ **: 3D Discrete gradient operator**
- μ**: Regularization parameter**

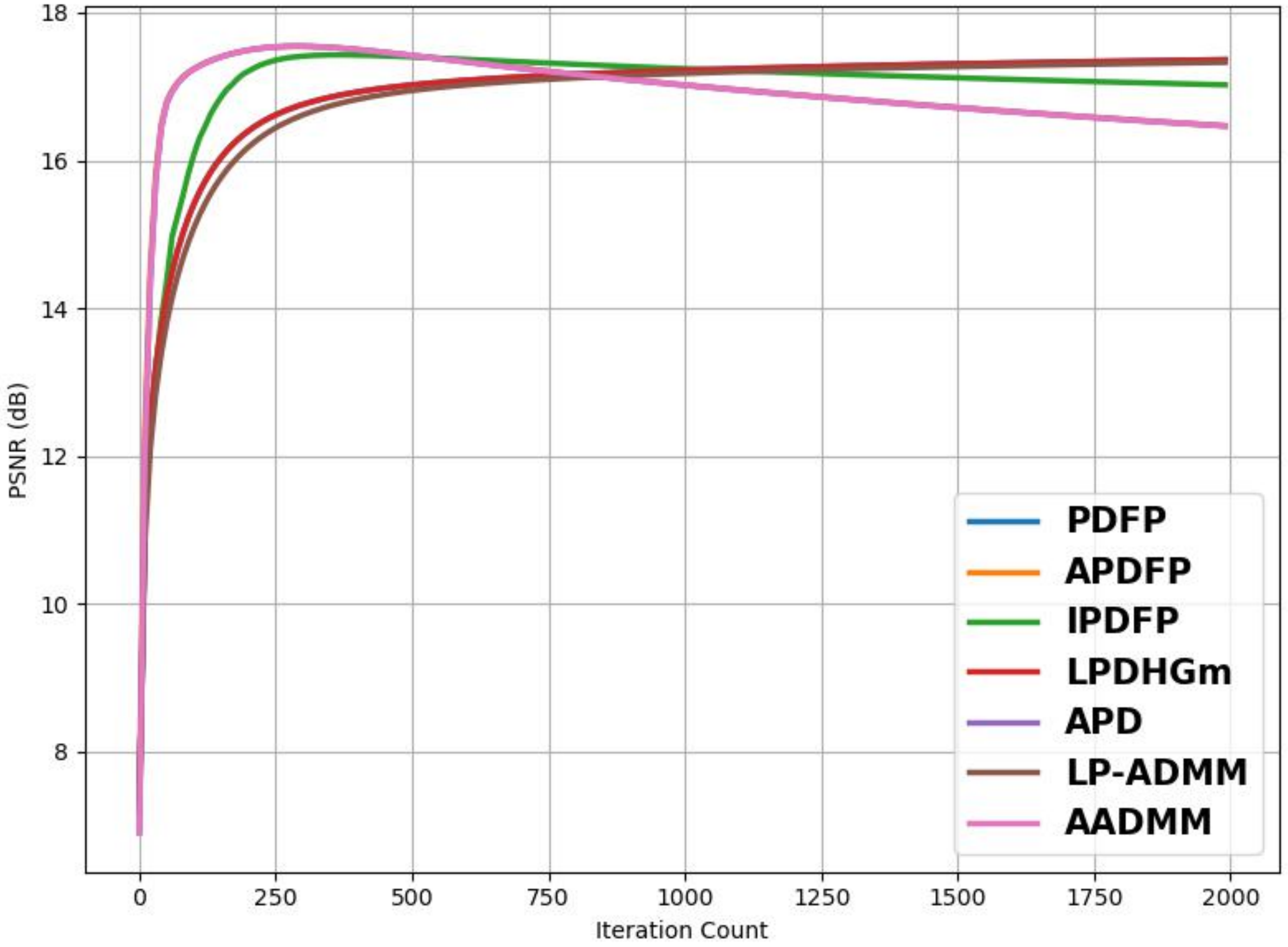
---

* H.Gao. "Fast parallel algorithms for the X-ray transform with cone beam geometry an its adjoint", Medical Physics (2012)

2D CT reconstruction



Objective v.s. iteration



PSNR v.s. iteration

2D CT reconstruction

Groud Truth

PDFP, PSNR=17.36

APDFP, PSNR=17.54

IPDFP PSNR=17.43

LPDHGm, PSNR=17.36

APD, PSNR=17.54

LP-ADMM, PSNR=17.39

AADMM, PSNR=17.54


Outline
Introduction
Applications
Mathematical Model
Algorithms Review
Motivation
Algorithm
APDFP
Convergence Results
Numerical Experiments
Conclusions

Conclusions

$$\min_{x \in \mathbb{R}^d} \; f(x) + g(x)$$

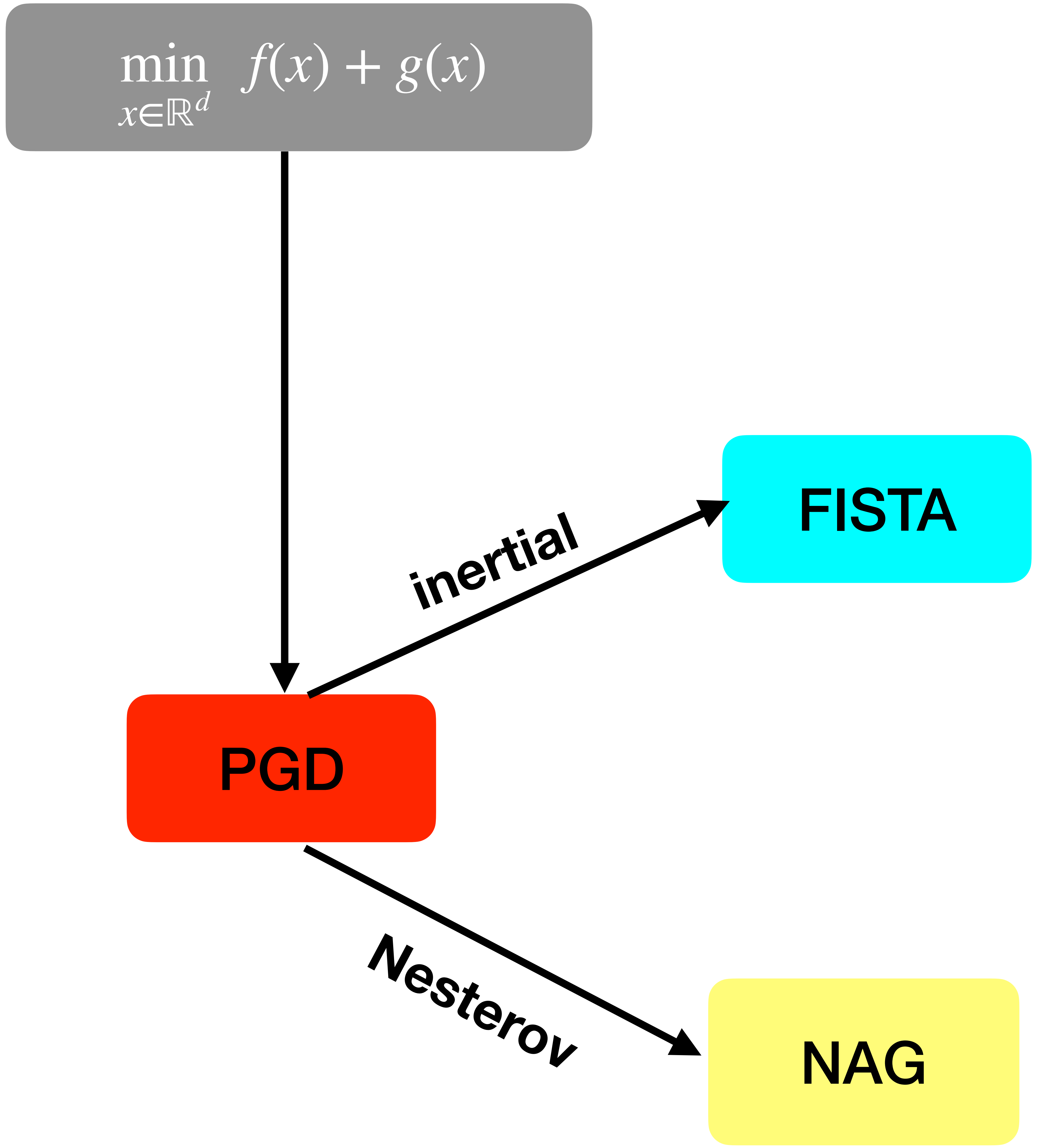
Conclusions

$$\min_{x \in \mathbb{R}^d} f(x) + g(x)$$

$\downarrow$

PGD

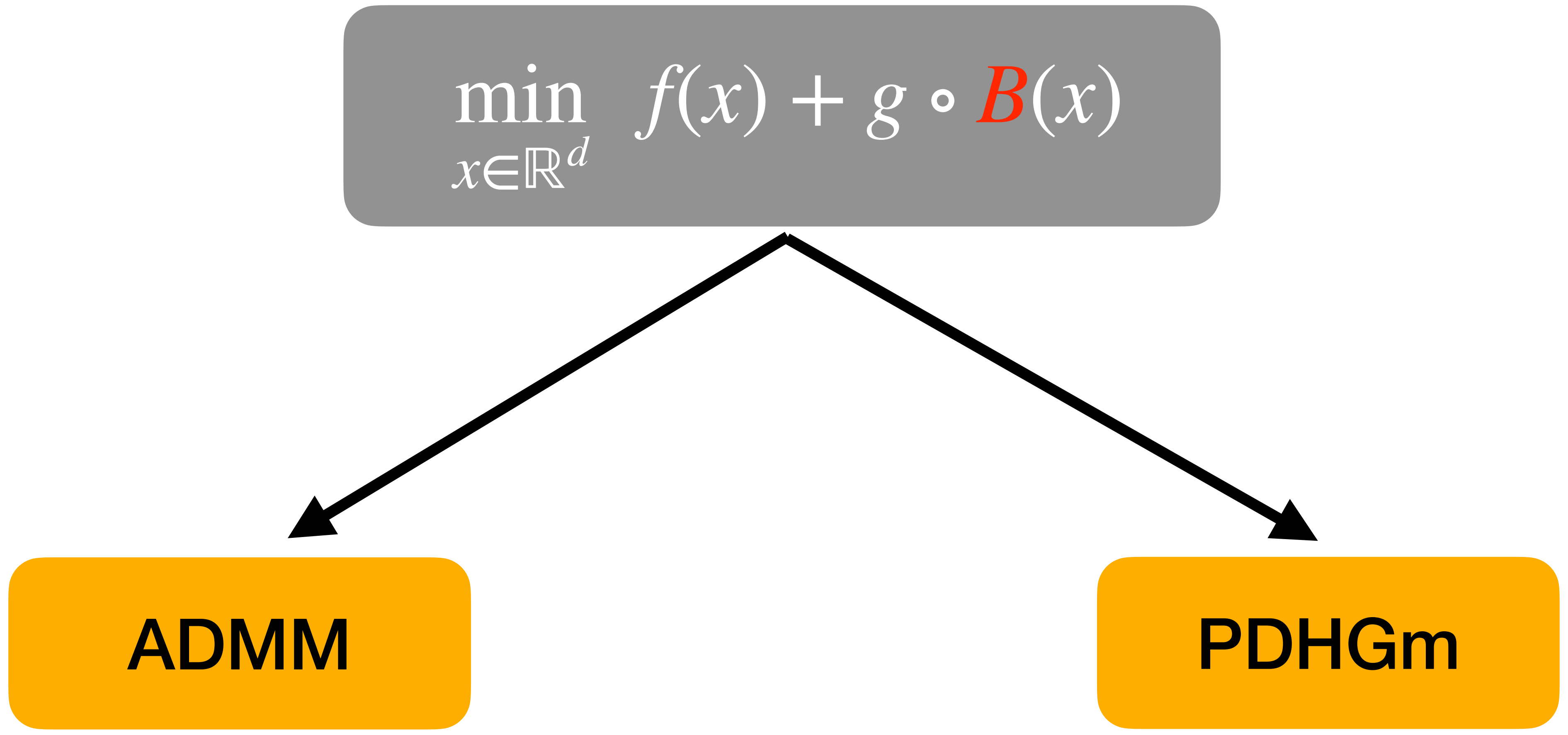
Conclusions



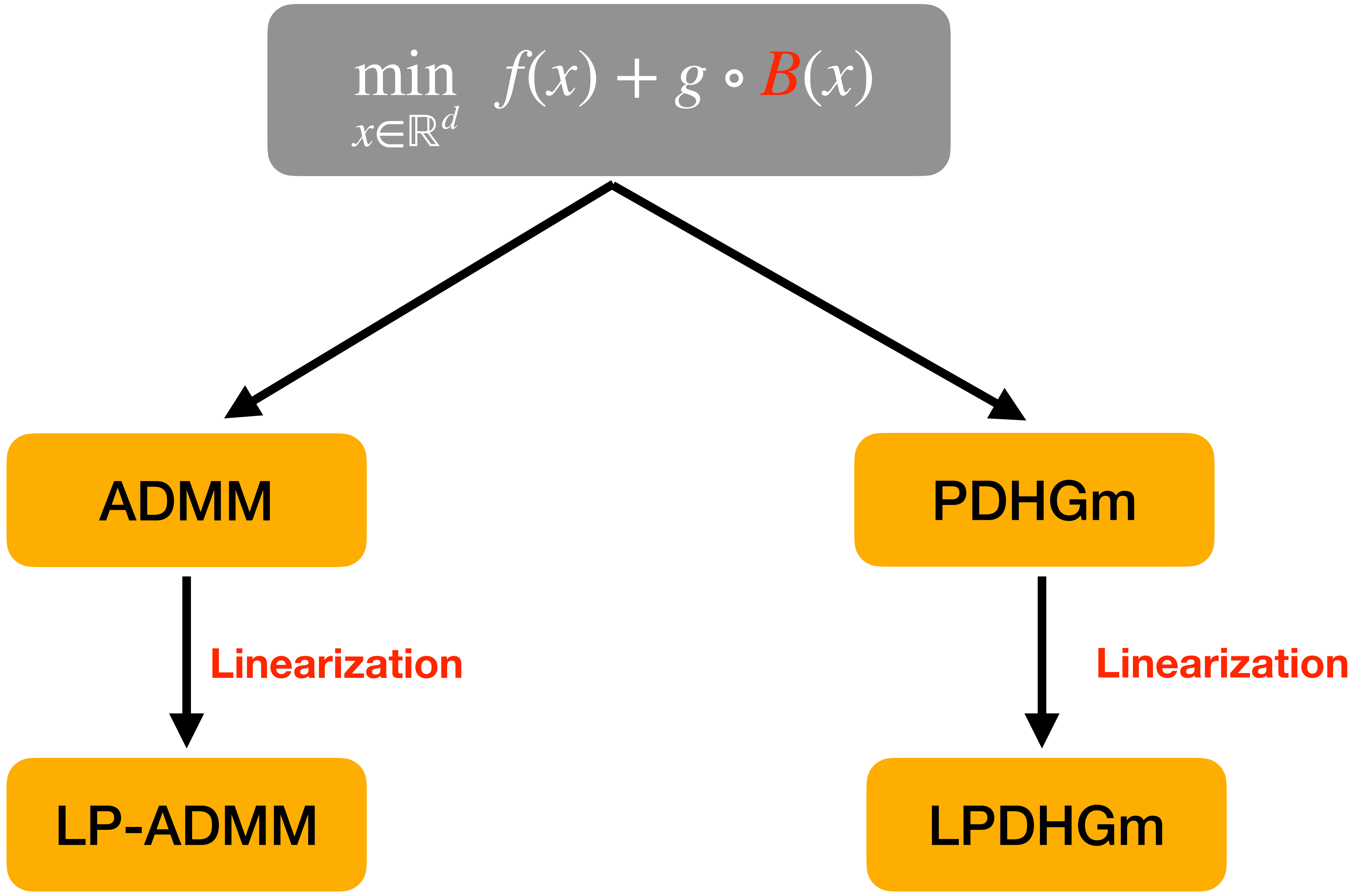
Conclusions

$$\min_{x \in \mathbb{R}^d} \; f(x) + g \circ B(x)$$

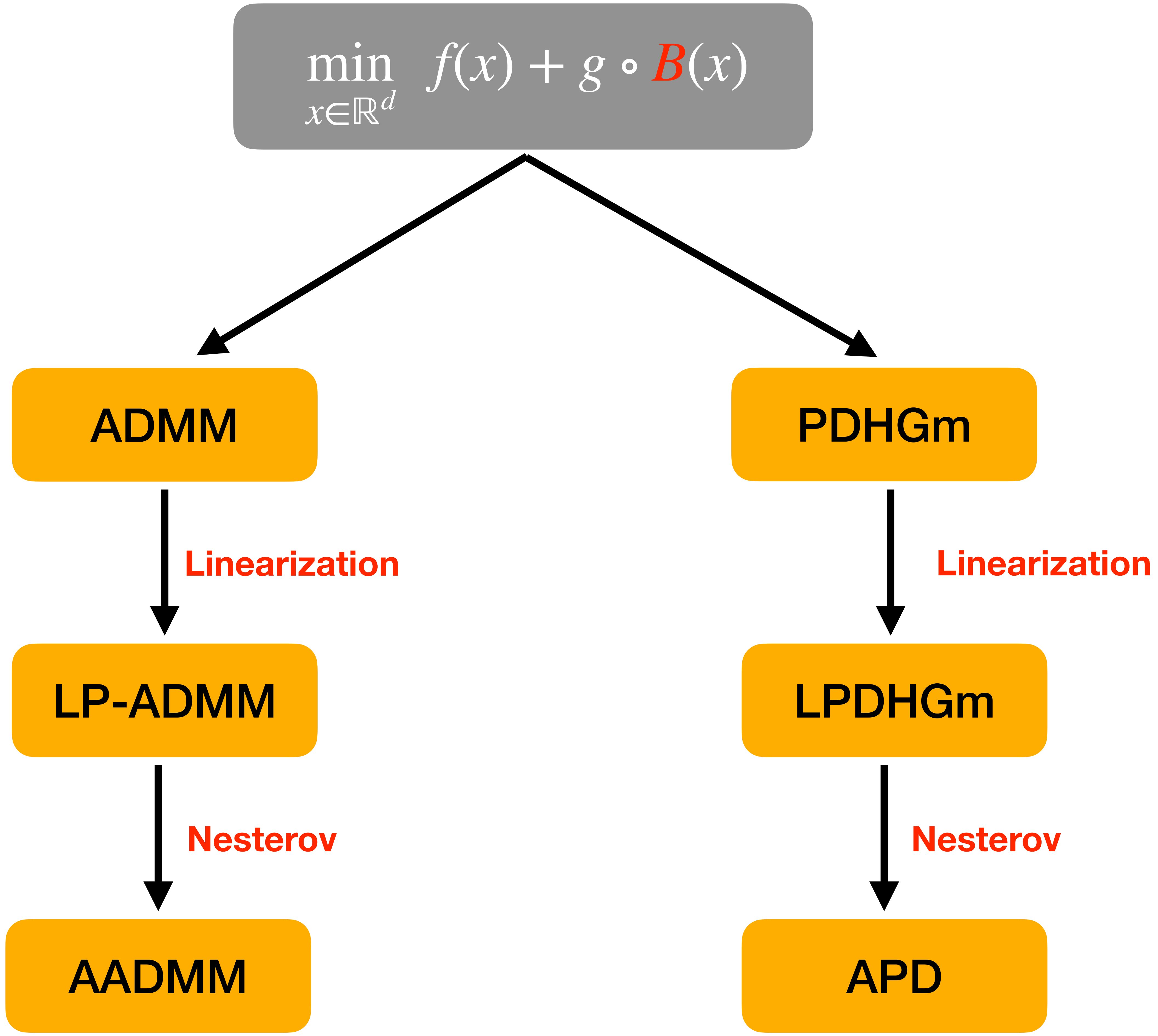
Conclusions



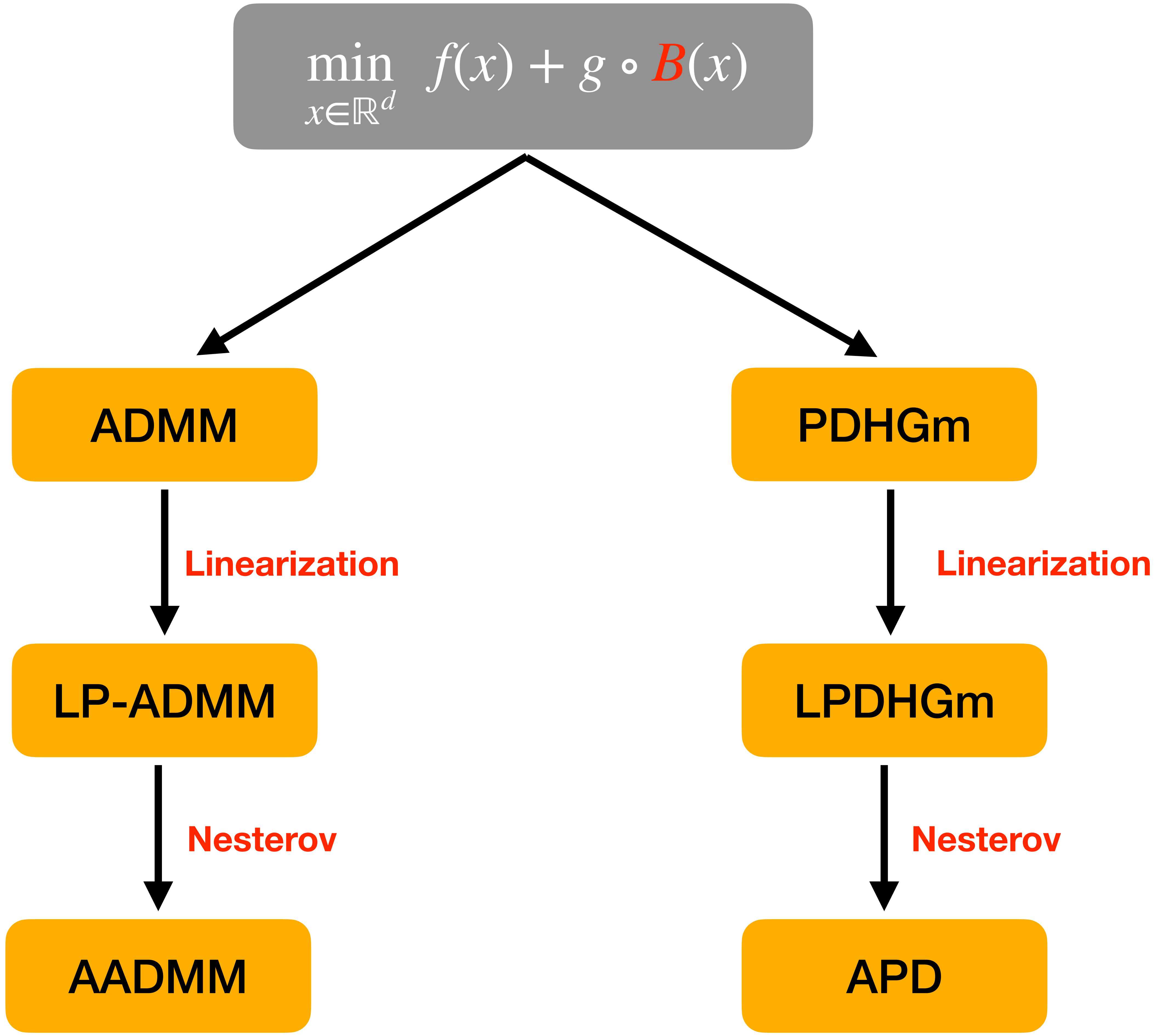
Conclusions



Conclusions

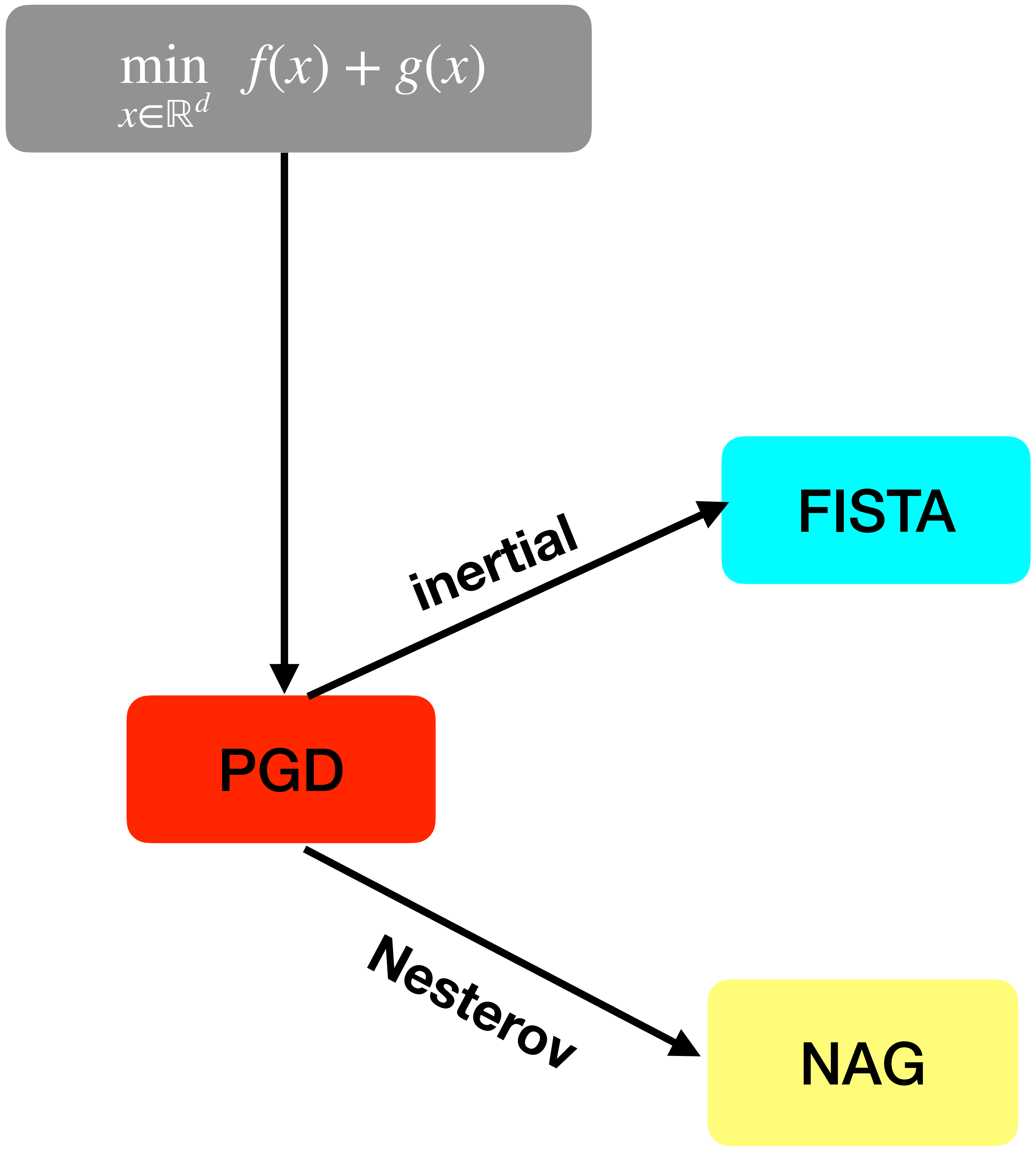


Conclusions



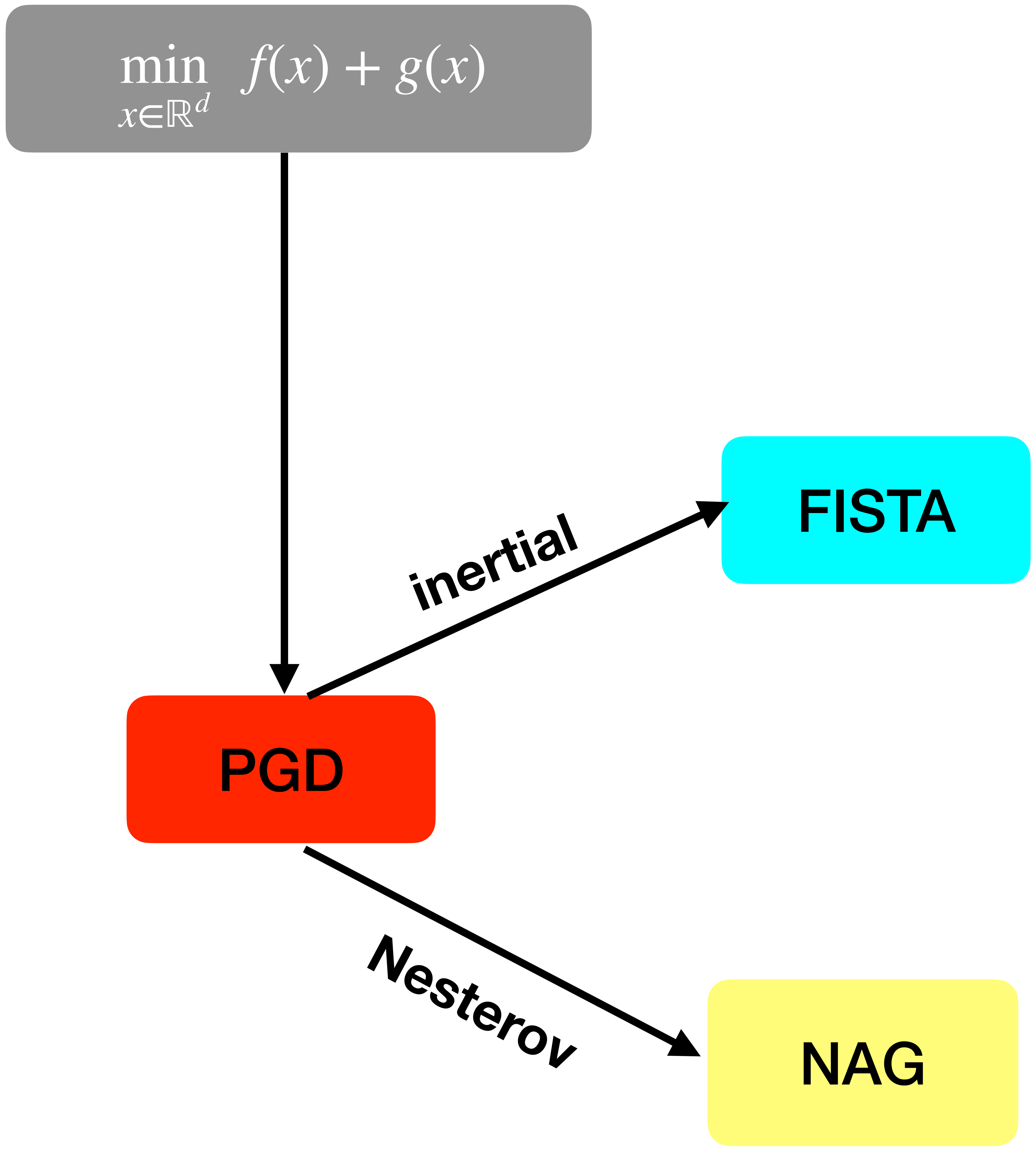
Parameter tuning are generally not easy

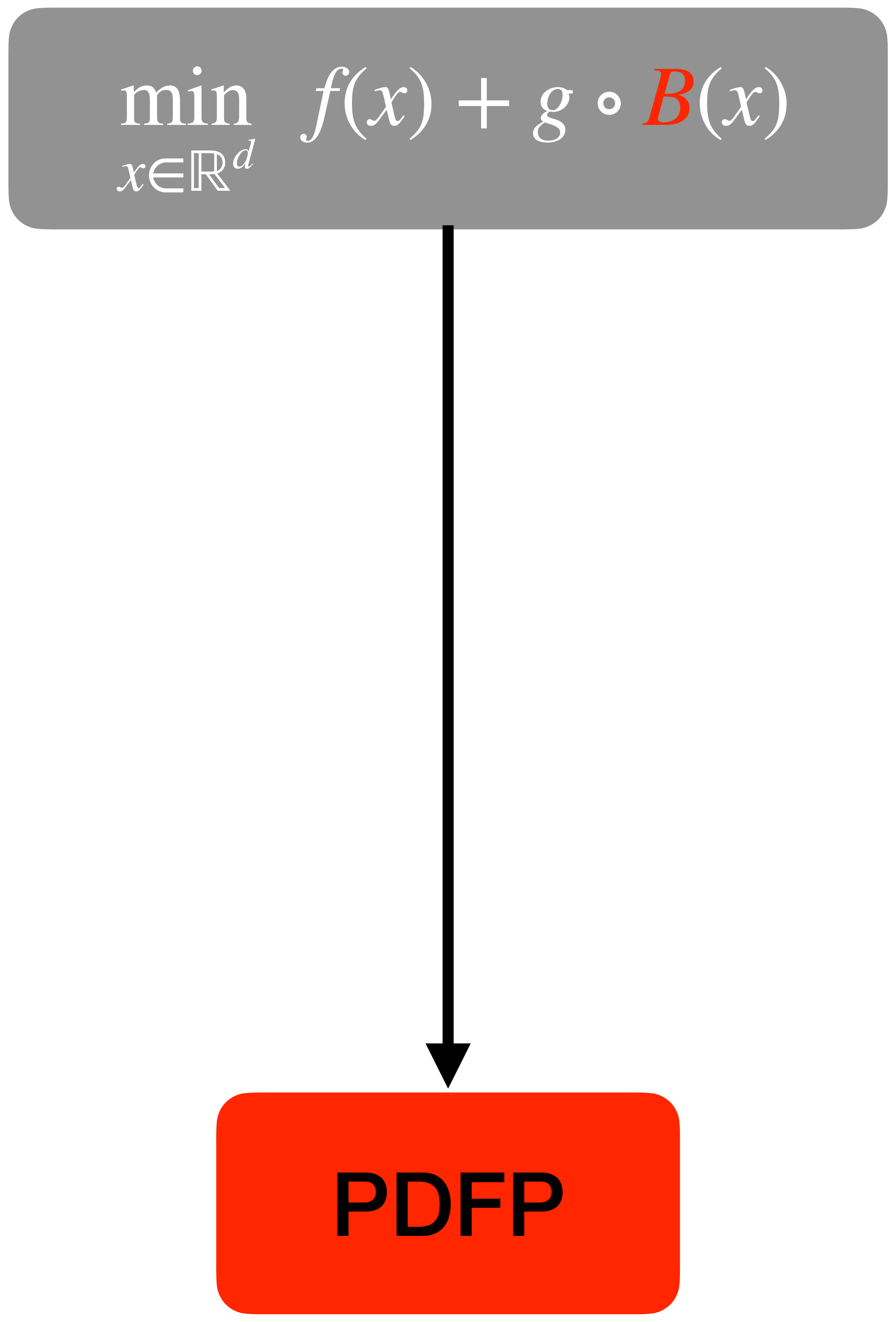
Conclusions



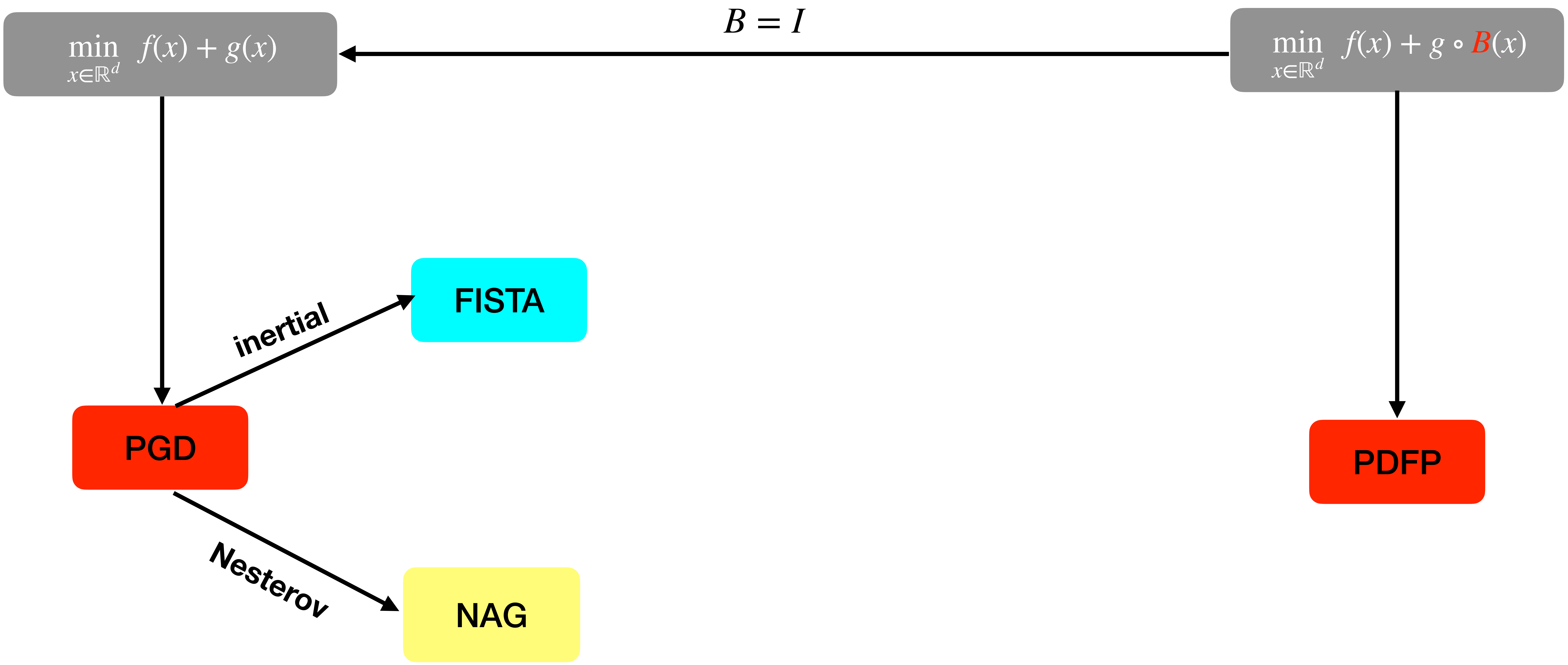
$$\min_{x \in \mathbb{R}^d} \; f(x) + g \circ B(x)$$

Conclusions

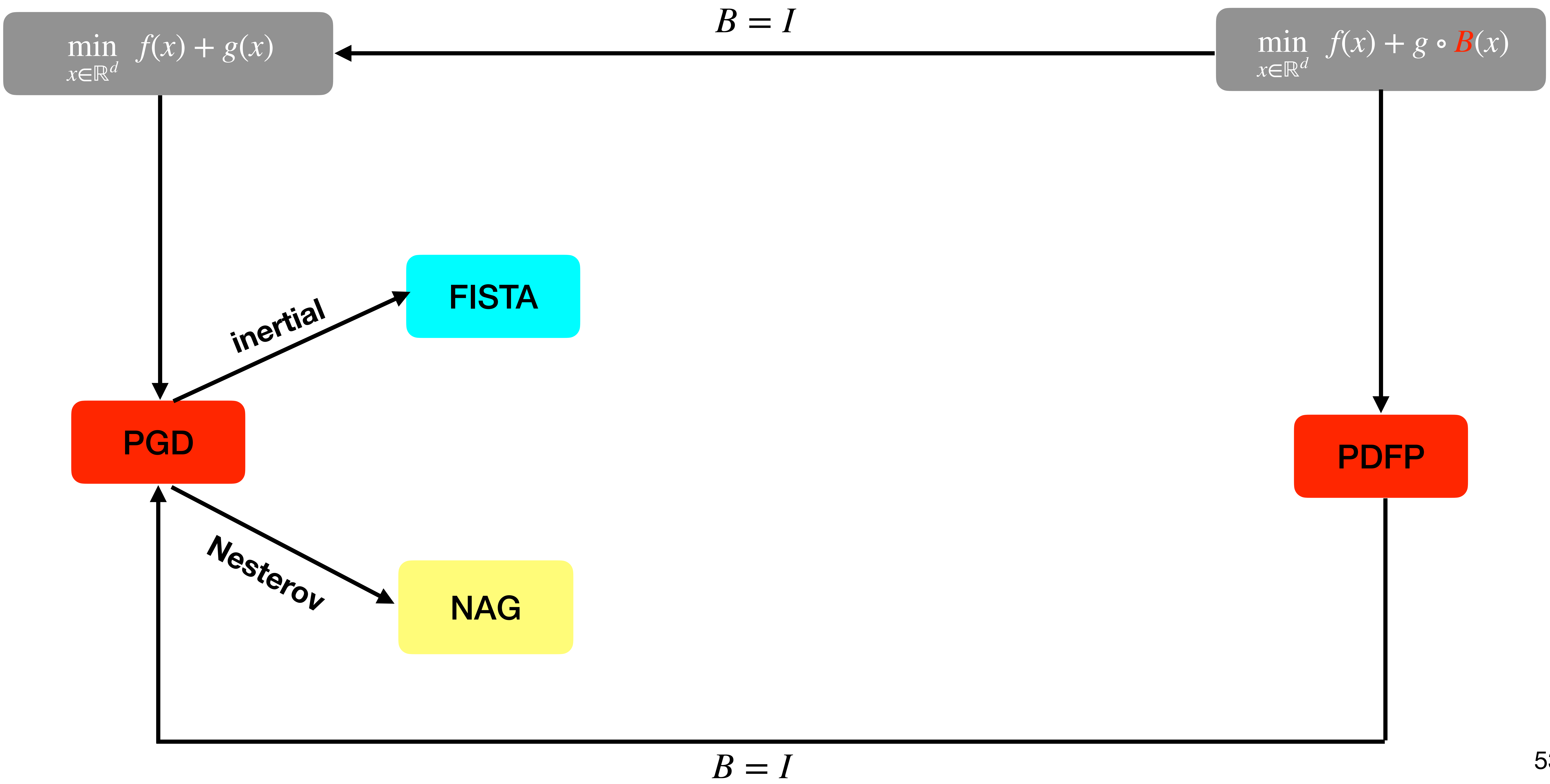




Conclusions



Conclusions



Conclusions

Conclusions

Conclusions

Conclusions

Conclusions

Thank You !

https://yananzhu6.github.io/Yanan.Zhu.github.io/talks/

Slide!