A taste of Primal-Dual with Alternating Projections and Optimal Transport

Mohammad Sadegh Salehi

March 16, 2023

University of Bath, Department of Mathematical Sciences, SAMBa CDT

Table of contents

Preliminaries

Preliminaries

Convex conjugate (Legendre–Fenchel conjugate):

- For $f : \mathcal{X} \to [-\infty, \infty]$

$$f^*(y) = \sup_{x \in \mathcal{X}} \langle y, x \rangle - f(x)$$

- **Biconjugate** f^{**} is largest lower semi-continuous ($\lim_{x \to x_0} \inf f(x) \geq f(x_0)$) convex function below f
- If f is l.s.c. and convex, then $f^{**} = f$ (a corollary of Hahn-Banach theorem)

Preliminaries

- Given a convex and l.s.c. $f : \mathcal{X} \to [-\infty, \infty]$ the **subgradient** at a point x is defined as

$$\partial f(x) := \{p \in \mathcal{X} : f(y) \geq f(x) + \langle p, y - x \rangle, \forall y \in \mathcal{X}\}$$

- For convex, proper, and l.s.c., **proximity** operator is

$$prox_{\tau f}(x) := \min_{y \in \mathcal{X}} f(y) + \frac{1}{2\tau} \|y - x\|^2$$

Fenchel–Rockafellar duality

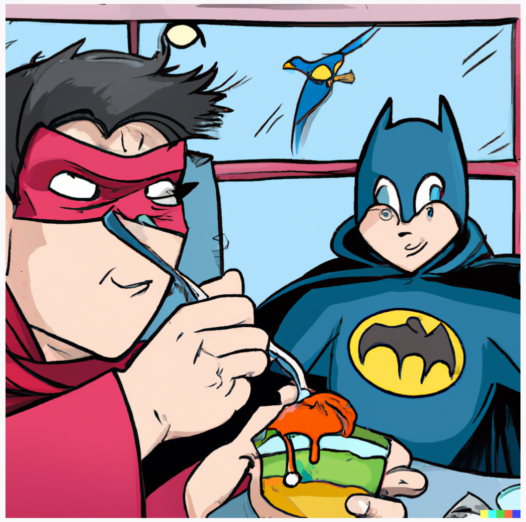
Definition

Let $f : \mathcal{Y} \to (-\infty, \infty]$ & $g : \mathcal{X} \to (-\infty, \infty]$ be convex and l.s.c., and $K : \mathcal{X} \to \mathcal{Y}$ be a linear operator. As $f = f^{**}$, we have

$$min_{x \in \mathcal{X}} f(Kx) + g(x) = \min_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \langle y, Kx \rangle - f^*(y) + g(x)$$

If $f(0) < \infty$ & g continuous at 0, or in **finite dimension** case $\exists x \in \mathcal{X}$ s.t. $Kx \in relint\{dom\ f\}$and $x \in relint\{dom\ g\}$,

$$\min_x \sup_y \langle y, Kx \rangle - f^*(y) + g(x) = \max_y \inf_x \langle y, Kx \rangle - f^*(y) + g(x)$$



$$= \max_y -f^*(y) - g^*(-K^* y)$$

Alternating Projections

Alternating Projections

Let $K_1, \dots, K_k \subset \mathbb{R}^N$ be convex sets which projection on each of them is simple (e.g., hyperplanes, halfspaces, etc).

Aim: Calculate the projection of $x \in \mathbb{R}^N$ onto the $\bigcap_{i=1}^{k} K_i$.

Problem

$$\min_{u \in \mathbb{R}^N} \frac{1}{2}\|x - u\|^2 + \sum_{i=1}^{k} \psi_i(u), \quad where \quad \psi_i(u) = \begin{cases} 0 & u \in K_i \\ +\infty & O.W. \end{cases}.$$

Historical Notes

- In 1949 Von Neumann (Neumann the Great) Proved the convergence in norm for two closed subsets of a Hilbert space
- In 1962, Halperin generalised Neumann's theorem for periodic update sequence (Using Kakutani's lemma)
- Convergence in finite dimension
- Convergence in the weak topology
- Not convergent in norm in infinite dimensional case with more than 2 closed sets

In our setting we concentrate on the closed and convex subsets. In this case the AP is convergent.

An example of Alternating Projections

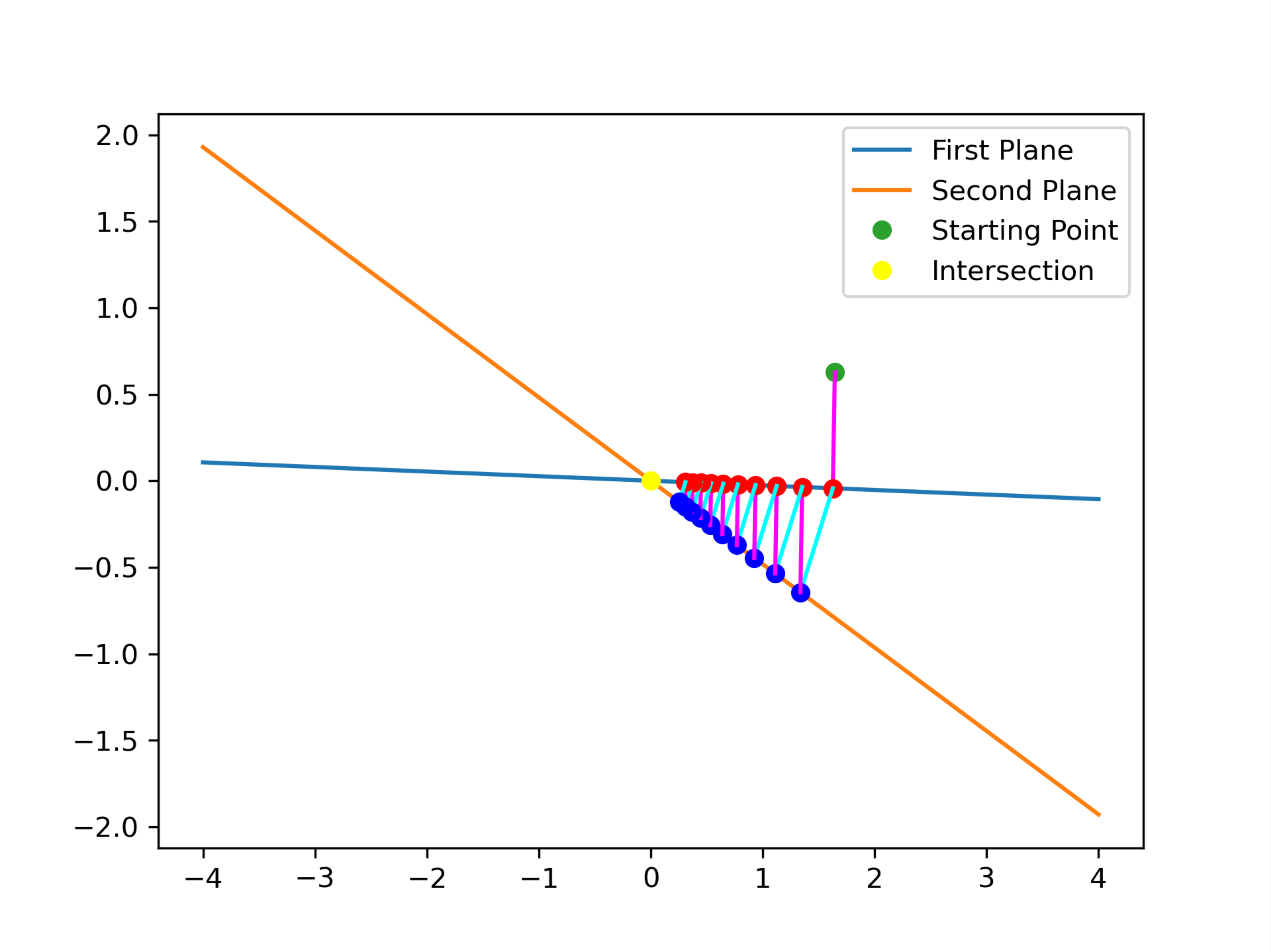


Figure 1: Alternating projections on two lines (hyperplanes) in $\mathbb{R}^2$

Dual problem

Problem

$$\min_{y\in\mathbb{R}^N} \frac{1}{2}\|x-y\|^2 + (\sum_{i=1}^{k}\psi_i)^*(y)$$

Note that $\bar{u}$ is the solution of the primal problem $\iff$ $\bar{y} = x - \bar{u}$ solves the dual problem.

Using inf-convolution $\left(\sum_{i=1}^{k}\psi_i\right)^*(y) = \inf\left\{\sum_{i=1}^{k}\psi_i^*(y_i) : \sum_{i=1}^{k} y_i = y\right\}$, the dual problem is

$$\inf_{(y_i)_{i=1}^{k}\in(\mathbb{R}^N)^k} \frac{1}{2}\left|x - \sum_{i=1}^{k} y_i\right|^2 + \sum_{i=1}^{k}\psi_i^*(y_i)$$

Dykstra

Using alternating minimisation on the dual problem, the main iteration of Dykstra's algorithm is

Dykstra iterations

$$\begin{cases} x_i^{n+1} = \Pi_{K_i}(x_{i-1}^n + y_i^n) \\ y_i^{n+1} = x_{i-1}^n + y_i^n - x_i^{n+1} \end{cases}$$

In 1985 Dykstra proved $x^n \xrightarrow{n\to\infty} \Pi_{\bigcap_{i=1}^k K_i}(x)$.

Accelerations

Anderson Acceleration on Dykstra

Algorithm 1 Anderson acceleration for Dykstra

Input: $x_0 \in \mathbb{R}^N,\ j \in \mathbb{N},\ \epsilon > 0$

Step 1: $i = 0$ and $x = x_0$

While $i \leq j$:

$$\begin{cases} x_i \leftarrow Dykstra(x) \\ x \leftarrow x_i,\ x_{\text{old}} = x \\ i \leftarrow i + 1 \end{cases}$$

Step 2: $U := [x_1 - x_0, \ldots, x_j - x_{j-1}]$

Step 3: Solve the linear system $(U^T U + \lambda I)z = \mathbf{1}$

$c := z / z^T \mathbf{1}$

Step 4: $x \leftarrow \sum_{k=0}^{j-1} c_k x_k$

If $\|x - x_{\text{old}}\| \geq \epsilon$:

$x_0 \leftarrow x$ then go to "step 1"

Else:

Output: x

Conjugate Gradient (CG)

Let convex sets be affine hyperplanes. For projection on these sets we have $\Pi_{ax=b}\, x_0 = x_0 + \left(\frac{b - a.x_0}{\|a\|}\right) a = (I - a \otimes a)x_0 + ba$. Then

$$x_1 = (\prod_{k=1}^{n}(I - a_k \otimes a_k))x_0 + (\prod_{k=2}^{n}(I - a_k \otimes a_k))b_1 a_1 + \cdots + (I - a_n \otimes a_n)b_{n-1}a_{n-1} + b_n a_n$$

We form a symmetric operator and the right-hand-side vector as follows:

$$A := x_0 - (M_1 \ldots M_n M_n \ldots M_1)x_0,$$

$$b := M_1 \ldots M_n M_n \ldots M_2 b_1 a_1 + \cdots + M_1 \ldots M_n b_n a_n + M_1 \ldots M_{n-1} b_n a_n + \cdots + M_1 b_2 a_2 + b_1 a_1$$

Finally, we apply CG on the linear system $Ax = b$ to find the desired point.

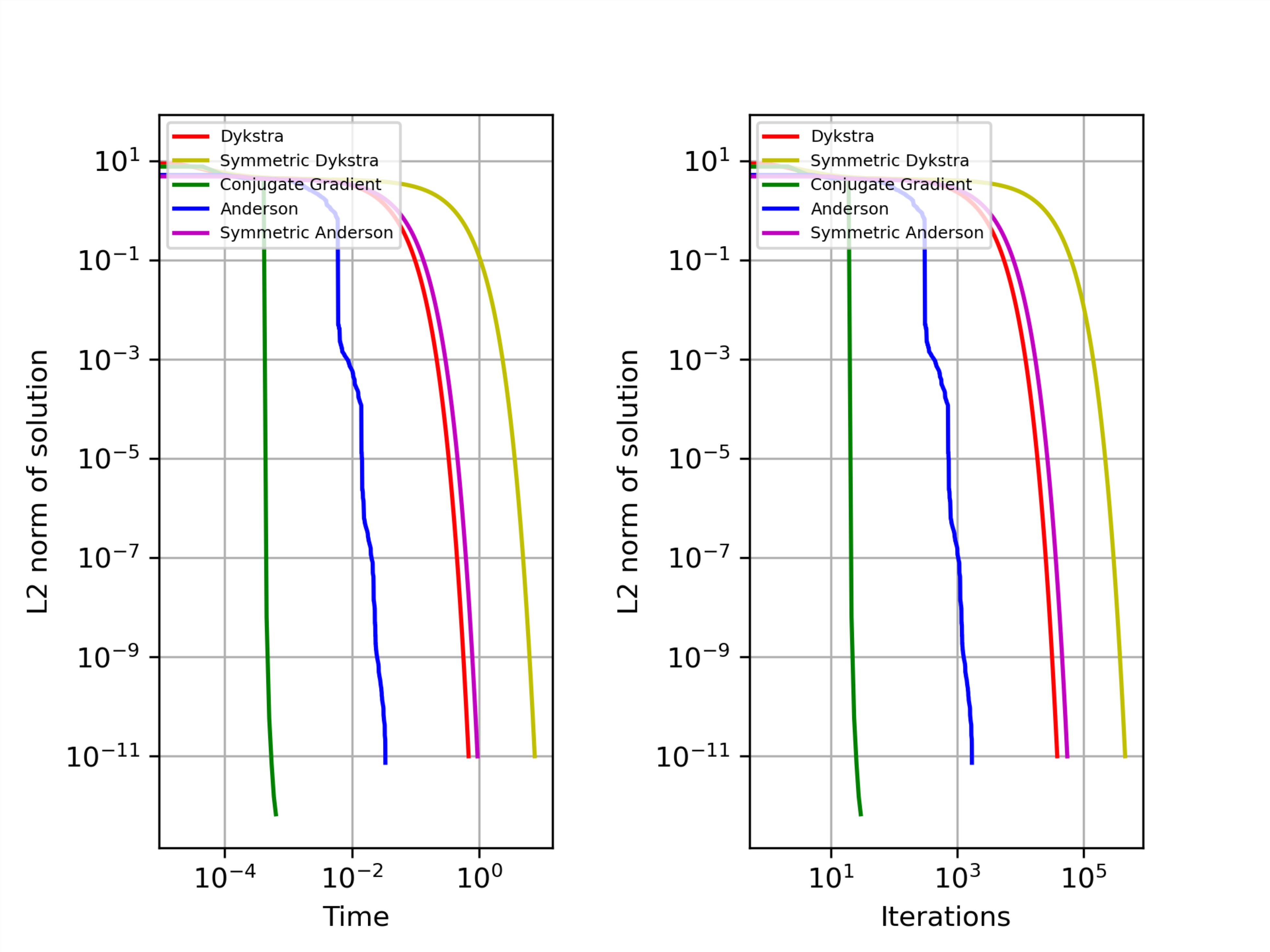


Figure 2: Projection of a random point on the intersection of 32 hyperplanes in $\mathbb{R}^{32}$ by setting $\epsilon = 10^{-11}$

Numerical Experiments

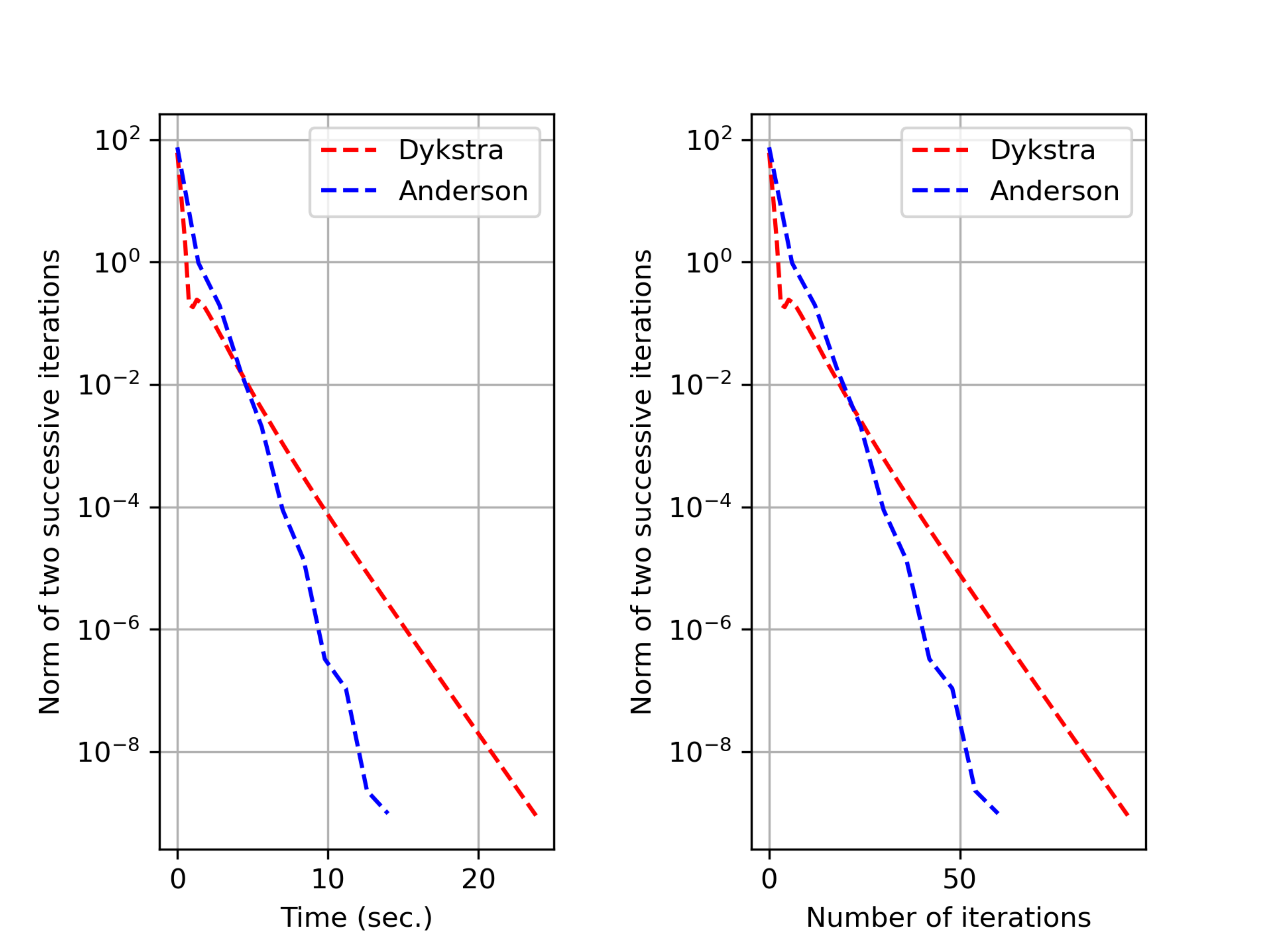


Figure 3: Projection of a random point on the intersection of 128 half-space in $\mathbb{R}^{128}$ by setting $\epsilon = 10^{-9}$

Applications

Schwarz method for solving Poisson equation

Let us consider the space $H = H_0^1(\Omega)$ and $\Omega = \Omega_1 \cup \Omega_2 \subset \mathbb{R}^2$. The subdomains are sufficiently smooth and the H is Hilbert

$$\langle u, v \rangle_H = \langle \nabla u, \nabla v \rangle_{L^2(\Omega)} = \int_\Omega \nabla u . \nabla v \, dx$$

Poisson equation with Dirichlet boundary condition

$$\begin{cases} \Delta u = f & in\ \Omega \\ u = 0 & on\ \Gamma \end{cases}$$

where $\Gamma = \partial\Omega$. Also, $\Gamma_k = \partial\Omega_k \cap \partial\Omega$ and $\gamma_k = \partial\Omega_k \backslash \partial\Omega$, $f \in L^2(\Omega)$

Goal: Finding a weak solution of the PDE above.

Idea: Alternating projections on a composite domain

Subproblems

By beginning from $u_0 \in H$ we obtain u_1 as a weak solution of

$$\begin{cases} \Delta u_1 = f & \textit{in } \Omega_1, \\ u_1 = 0 & \textit{on } \Gamma_1, \\ u_1 = u_0 & \textit{on } \gamma_1 \end{cases}$$

After finding u_1 by setting $u_1 = u_0$ on $\Omega_2 \backslash \Omega_1$ we extend u_1 to Ω. Then we obtain u_2 by solving the following problem

$$\begin{cases} \Delta u_2 = f & \textit{in } \Omega_2, \\ u_2 = 0 & \textit{on } \Gamma_2, \\ u_2 = u_1 & \textit{on } \gamma_2 \end{cases}$$

Let $Y_k = H_0^1(\Omega_k)$, $M_k = Y_k^{\perp}$ and Π_k be orthogonal projection onto M_k for $k \in \{1, 2\}$. Each Y_k can be considered as a close subspace of H by extending functions defined on Ω_k by zero to whole Ω. Also $M = M_1 \cap M_2$.

Convergence

$$u - u_0 = \underbrace{(u - u_1)}_{\in M_1} + \underbrace{(u_1 - u_0)}_{\in M_1^\perp} \implies \Pi_1(u - u_0) = u - u_1$$

Similarly $\Pi_2(u - u_1) = u - u_2$ and this way continues. For each $n \geq 1$, $x_n := u - u_n$. Thus,

$$x_{2n} = (\Pi_2 \Pi_1)^n x_0, n \geq 1$$

By von Neumann's theorem,

$$\|x_{2n} - \Pi_M x_0\| \xrightarrow{n\to\infty} 0$$

$$\implies \|x_{2n+1} - \Pi_M x_0\| = \|\Pi_1(x_{2n} - \Pi_M x_0)\| \leq \|x_{2n} - \Pi_M x_0\| \xrightarrow{n\to\infty} 0$$

$$\implies \|x_n - \Pi_M x_0\| \xrightarrow{n\to\infty} 0$$

Since $M_1 \cap M_2 = Y_1^\perp \cap Y_2^\perp = (Y_1 + Y_2)^\perp$ and the subspace $Y = Y_1 + Y_2$ is dense in H, we have

$$M = Y^\perp = \{0\} \implies x_n \xrightarrow{n\to\infty} 0 \implies \|u_n - u\| \xrightarrow{n\to\infty} 0$$

Numerical example

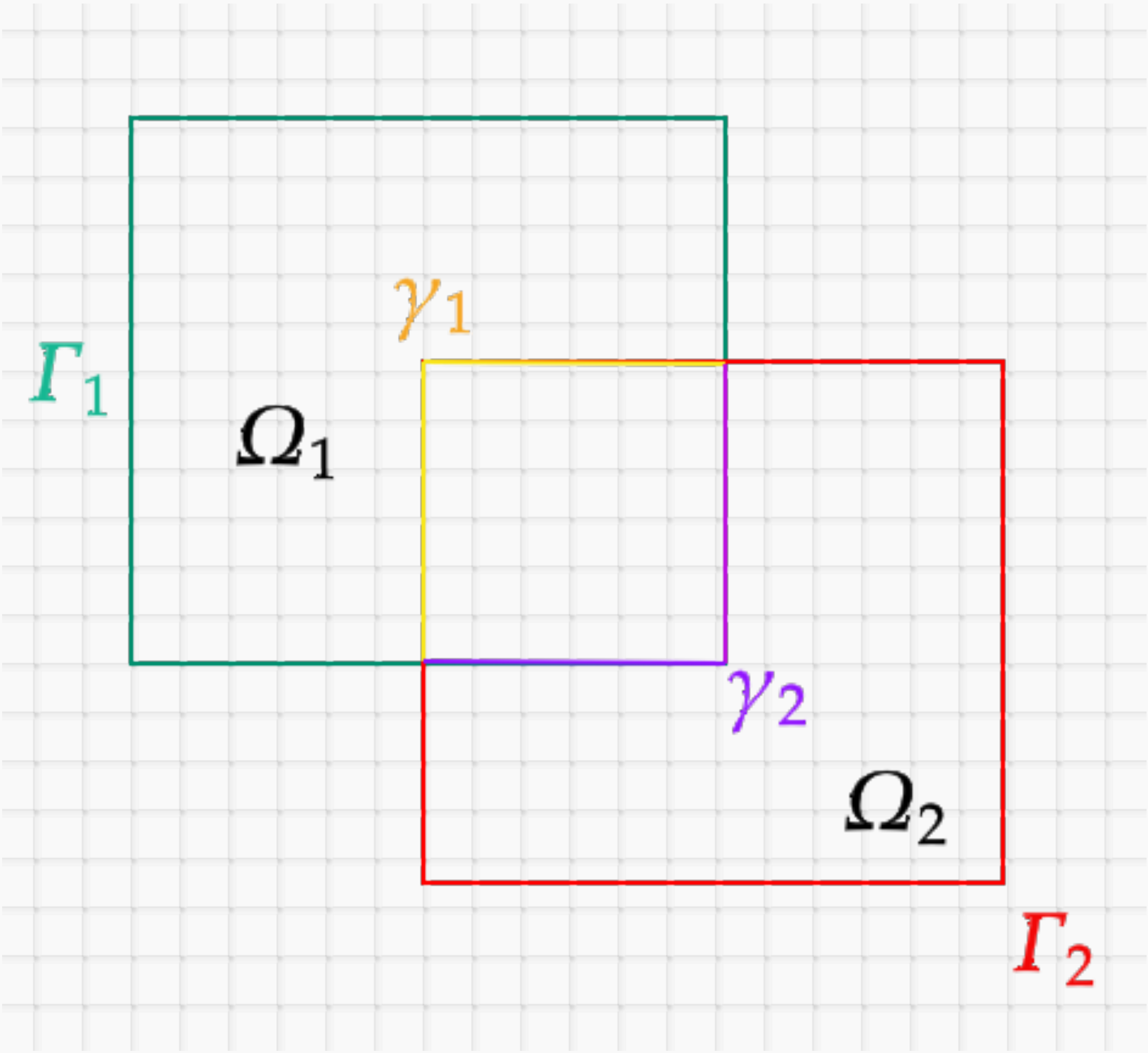


Figure 4: Composite Domain

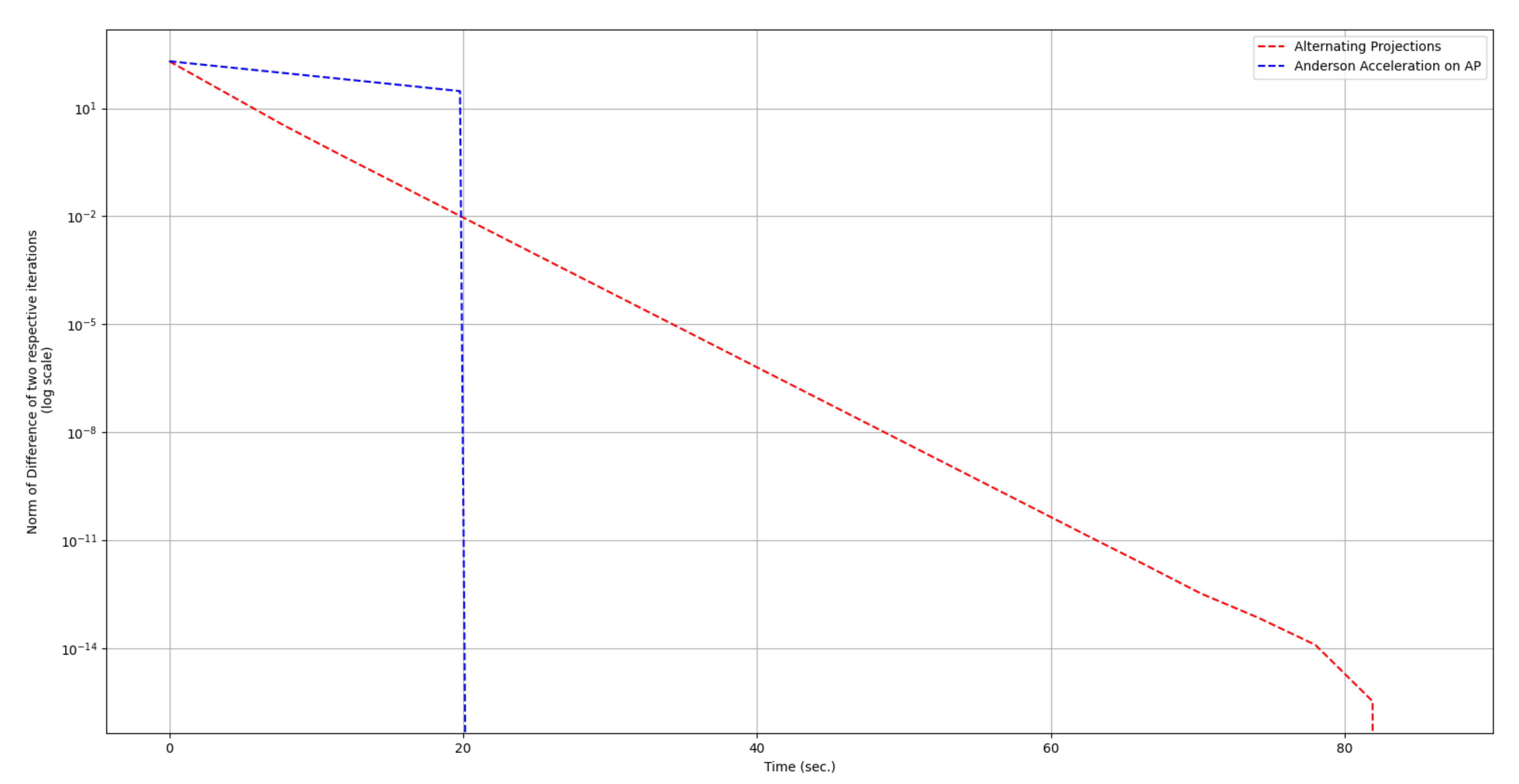


Figure 5: Anderson acceleration on Schwarz method for solving Poisson equation

Some other Applications

- Linear Classification
- SDP feasibility & Special cases of matrix completion
- Mutual applications with coordinate descent for regularised regression as an equivalent method
- Mutual applications with ADMM in case we have 2 as an equivalent method

Note that the randomised versions of Dykstra by stochastic coordinate descent on the dual variables exist that under some assumptions provide interesting results.

Optimal Transport (OT)

Setting

Consider the optimal transport problem

$$\min_{X \in \Sigma_{\mu\nu}} \langle C, X \rangle, \qquad \Sigma_{\mu\nu} = \{X \in \mathbb{R}_+^{n\times n} : X\mathbf{1} = \mu, X^T\mathbf{1} = \nu\},$$

where

- $\mathbb{1} = (1, 1, \ldots, 1)\ \mathbb{R}^n$
- $C \in \mathbb{R}_+^{n\times n}$ is a given *cost matrix*
- $\Delta^n = \{x \in \mathbb{R}^N \ :\ \sum_{i=1}^n x_i = 1\ ,\ x_i \geq 0 \textit{ for } i = 1, \ldots, n\}$ is unit simplex
- $\mu \in \Delta^n, \nu \in \Delta^n$

Hybrid Primal Dual

$$\min_{x \in X} \max_{y \in Y} \mathcal{L}(x, y) = \langle Kx, y \rangle + f(x) + g(x) - h^*(y)$$

$$\begin{cases} K = I \text{ (Identity matrix)} \\ f(Y) = 0 \\ g(Y) = \underset{\{0\}}{\delta}(Y) \text{ (Strongly convex)} \\ h^*(X) = \underset{\Sigma_{\mu\nu}}{\delta}(X) + \langle C, X \rangle \end{cases}$$

$$\min_{X} \max_{Y} \langle C, X \rangle + \underset{\Sigma_{\mu\nu}}{\delta}(X) + X : Y - \delta_{\{0\}}(Y)$$

HPD Method Iteration

For $x^0, \bar{x}^0 \in \mathrm{dom}\xi_{\mathcal{X}}$, $y_0 \in \mathrm{dom}\xi_Y$, and given nonnegative sequences $\{\tau_k\}_k$, $\{\sigma_k\}_k$, $\{\theta_k\}_k$:

$$
\begin{aligned}
y_{k+1} &= \operatorname*{argmin}_{y \in \mathcal{Y}} h^*(y) - \langle K\bar{x}^k, y\rangle + \frac{1}{\sigma_k} D_{\mathcal{Y}}(y, y_k) \\
x^{k+1} &= \operatorname*{argmin}_{x \in \mathcal{X}} g(x) + \langle Kx, y_k\rangle + \frac{1}{\tau_k} D_{\mathcal{X}}(x, x^k) \\
\bar{x}^{k+1} &= x^{k+1} + \theta_k(x^{k+1} - x^k).
\end{aligned}
$$

In case $\tau_k \equiv \tau_0$ and $\sigma_k \equiv \sigma_0$ are constant, taking $\theta_k \equiv 1$ and $\tau_0\sigma_0 L^2 \leq 1$, we have for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$,

$$
\mathcal{L}_{x,y}(\hat{x}^N, \hat{y}_N) \leq \frac{1}{T_N}\left(\frac{1}{\tau_0} D_{\mathcal{X}}(x, x^0) + \frac{1}{\sigma_0} D_{\mathcal{Y}}(y, y_0)\right)
$$

where $T_N = \frac{N}{2},\ x^N = \frac{1}{N}\sum_{k=1}^N x_k,\ y_N = \frac{1}{N}\sum_{k=1}^N y_k.$

Bound of duality gap

Assuming ϵ type-2 error in calculating proximal operator and having an error $\|e_k\|$ at each iteration (typical in algorithms like Sinkhorn), we have can derive the following bound

$$\langle C, \bar{X} - X^* \rangle \leq \frac{1}{T_N \sigma_0}(1 + (2N\sqrt{2} - T_N \sigma_0 \|C\|)\epsilon),$$

where $\|e_k\| \leq \epsilon$.

- By applying Nesterov acceleration on the update of X in the coordinate descent, we can derive and accelerated algorithm.
- Choosing the acceleration parameter is quite challenging.
- The stopping criterion would be similar to Sinkhorn and Round $\|\mu - X_{new}\mathbf{1}\|_1 + \|\nu - X_{new}^T\mathbf{1}\|_1 < \ error_{max}$
- Combining *HPD* and Round, Chambolle et al., 2023 proposed a tighter bound and introduced accelerated method with backtracking.

Summary

Summary

We talked about

- Fundamental definitions and theorems in Convex Analysis
- Alternating Projections
- Primal-Dual Alternating Projections
- Acceleration Methods
- An application to solving PDEs
- Discrete Optimal Transport, HPD setting, acceleration, and bound of Duality Gap

I would love to answer your questions :)

Thank you!

With Primal-Dual man

References

- Antonin Chambolle, Daniel Cremers, Thomas Pock. “A Convex Approach to Minimal Partitions.” 2011. ⟨hal-00630947⟩
- J. P. Boyle, R. L. Dykstra, “A method for finding projections onto the intersection of convex sets in Hilbert spaces”, in Advances in Order Restricted Statistical Inference (Iowa City, IA, 1985), Lecture Notes in Statist. 37, Springer, Berlin, 1986, pp. 28–47
- Damien Scieur, Alexandre d’Aspremont, Francis Bach. “Regularized Nonlinear Acceleration.” Mathematical Programming, Springer Verlag, 2018, ⟨10.1007/s10107-018-1319-8⟩. ⟨hal-01384682v2⟩
- Rasch, J., Chambolle, A. “Inexact first-order primal–dual algorithms.” Comput Optim Appl 76, 381–430 (2020). https://doi.org/10.1007/s10589-020-00186-y
- Deutsch, F. (1992) The Method of Alternating Orthogonal Projections. In: Singh S.P. (eds) Approximation Theory, Spline

References

Netyanun, A., & Solmon, D. (2006). Iterated Products of Projections in Hilbert Space. The American Mathematical Monthly, 113(7), 644-648. doi:10.2307/27642008

Nesterov, Y. (1983) A Method for Solving a Convex Programming Problem with Convergence Rate O(1/K2). Soviet Mathematics Doklady, 27, 372-367.

Von Neumann, J. “On rings of operators. Reduction theory”. Annals of Mathematics, pp. 401–485, 1949.

Halperin, I. “The product of projection operators”. Acta Sci.Math.(Szeged), vol. 23, no. 1, pp. 96–99, 1962.

References

Laurent, P., Legendre, G. & Salomon, J. On the method of reflections. Numer. Math. 148, 449–493 (2021). https://doi.org/10.1007/s00211-021-01207-6

Badea, C., Grivaux, S., and Müller,V. "The rate of convergence in the method of alternating projections," St. Petersburg Mathematical Journal, vol. 23, no. 3, pp. 413–434, 2012.

Lions, P. L. "On the Schwarz alternating method. I," in First international symposium on domain decomposition methods for partial differential equations, 1988, pp. 1–42.

Anderson, Donald G. "Iterative Procedures for Nonlinear Integral Equations." Journal of the ACM, 12(4):547–560, 1965. https://doi.org/10.1145/321296.321305