F. Proof of Arbitrary Initialization

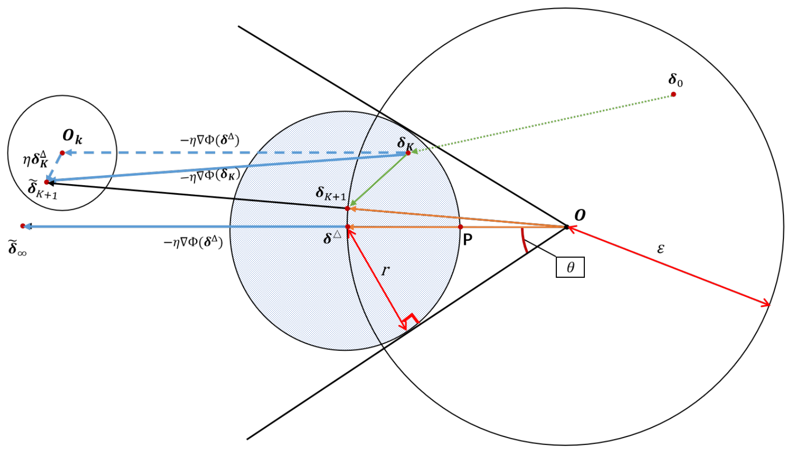

Figure F1. The process of PGD: $\boldsymbol{\delta}_K \to \tilde{\boldsymbol{\delta}}_{K+1} \to \boldsymbol{\delta}_{K+1}$. The Light blue filled ball represents $\mathbb{B}(\boldsymbol{\delta}^{\triangle}, r)$. The green trajectory of $\boldsymbol{\delta}_k$ is generated by PGD. The solid blue arrow means $-\eta\nabla\Phi(\boldsymbol{\delta}_K)$. The dotted blue arrows indicate auxiliary vectors $-\eta\nabla\Phi(\boldsymbol{\delta}^{\triangle})$ and $\eta\boldsymbol{\delta}_k^{\triangle}$, where $\eta < \frac{1}{2M}$ and $\boldsymbol{\delta}_k^{\triangle} := \nabla\Phi(\boldsymbol{\delta}^{\triangle}) - \nabla\Phi(\boldsymbol{\delta}_k)$ with $\|\boldsymbol{\delta}_k^{\triangle}\| \le Mr$ (smooth).

Theorem 3.1 shows that there exists $\theta > 0$ such that if the initialization is chosen from $\mathbb{S}(\mathbf{0}, \epsilon) \cap \mathbb{B}(\boldsymbol{\delta}^{\triangle}, 2\epsilon \sin\frac{\theta}{2})$, then PGD exhibits local linear convergence. We now show that PGD achieves global convergence and eventually at least local linear convergence for general initialization.

When the initial point $\boldsymbol{\delta}_0$ is randomly chosen, prior work (Nesterov, 2018; Boyd, 2014; Bubeck, 2015) shows that the sequence $\boldsymbol{\delta}_k$ generated by PGD converges to the optimal point $\boldsymbol{\delta}^{\triangle}$ at a global sublinear rate. Therefore, for any $r > 0$, there exists $K(r) \in \mathbb{Z}^+$ such that $\boldsymbol{\delta}_{K+i} \in \mathbb{B}(\boldsymbol{\delta}^{\triangle}, r) \cap \mathbb{B}(\mathbf{0}, \epsilon)$ for all $i \in \mathbb{Z}^+$, where this process is shown in Figure F1.

Next, we specify a suitable choice of r to ensure that $\boldsymbol{\delta}_{K+1}$ satisfies the initialization condition of Theorem 3.1, which guarantees that the sequence $\boldsymbol{\delta}_{K+i}$ for $i = 1, 2, \ldots$ converges at a linear rate.

1) To ensure $\tilde{\boldsymbol{\delta}}_{K+1} \notin \mathbb{B}(\mathbf{0}, \epsilon)$, in the worst case (when $\boldsymbol{\delta}_K$ is moved to point P), we require:

$$\epsilon - r + \eta\|\nabla\Phi(\boldsymbol{\delta}^{\triangle})\| - \eta\|\boldsymbol{\delta}_K^{\triangle}\| > \epsilon,$$

which leads to the sufficient condition:

$$\epsilon - r + \eta\|\nabla\Phi(\boldsymbol{\delta}^{\triangle})\| - Mr\eta > \epsilon \Rightarrow r < \frac{\eta\|\nabla\Phi(\boldsymbol{\delta}^{\triangle})\|}{M\eta + 1}. \tag{1}$$

2) To ensure $\tilde{\boldsymbol{\delta}}_{K+1} \in \left\{\tilde{\boldsymbol{\delta}} \mid 0 \le \angle(\tilde{\boldsymbol{\delta}}, \boldsymbol{\delta}^{\triangle}) < \theta\right\}$, since $\eta\|\boldsymbol{\delta}_k^{\triangle}\| \le \eta Mr < r$, it is sufficient that:

$$r < \epsilon \sin\theta.$$

3) We also require $\tilde{\boldsymbol{\delta}}_{K+1} \notin \mathbb{B}(\boldsymbol{\delta}^{\triangle}, r)$, whose sufficient condition is:

$$\eta\|\nabla\Phi(\boldsymbol{\delta}_K)\| > \eta(\|\nabla\Phi(\boldsymbol{\delta}^{\triangle})\| - Mr) > 2r \Rightarrow r < \frac{\eta\|\nabla\Phi(\boldsymbol{\delta}^{\triangle})\|}{M\eta + 2} \tag{2}$$

Conclusion: For any

$$r < \min\left\{\epsilon\sin\theta, \frac{\eta\|\nabla\Phi(\boldsymbol{\delta}^{\triangle})\|}{M\eta + 2}\right\},$$

there exists $K \in \mathbb{Z}^+$ such that

$$\tilde{\boldsymbol{\delta}}_{K+1} \in \left\{\tilde{\boldsymbol{\delta}} \mid 0 \le \angle(\tilde{\boldsymbol{\delta}}, \boldsymbol{\delta}^{\triangle}) < \theta,\ \tilde{\boldsymbol{\delta}} \notin \mathbb{B}(\mathbf{0}, \epsilon)\right\}.$$

If $\boldsymbol{\delta}_{K+1} \neq \boldsymbol{\delta}^{\triangle}$, then it satisfies the initialization condition of Theorem 3.1, and the subsequent sequence $\boldsymbol{\delta}_{K+i}$ for $i = 1, 2, \ldots$ converges at least at a linear rate.

If $\boldsymbol{\delta}_{K+1} = \boldsymbol{\delta}^{\triangle}$, the sequence $\boldsymbol{\delta}_k$ exhibits finite-step convergence. Prior works [1–5] show that such convergence is effectively instantaneous and can be considered faster than linear (See Figure F2). This does not contradict our conclusion, as our paper emphasizes that PGD converges at least at a linear rate, and the lower bound applies under the specific initialization assumed in Theorem 3.1.

In summary, PGD achieves global convergence and eventually at least local linear convergence even under general initialization.

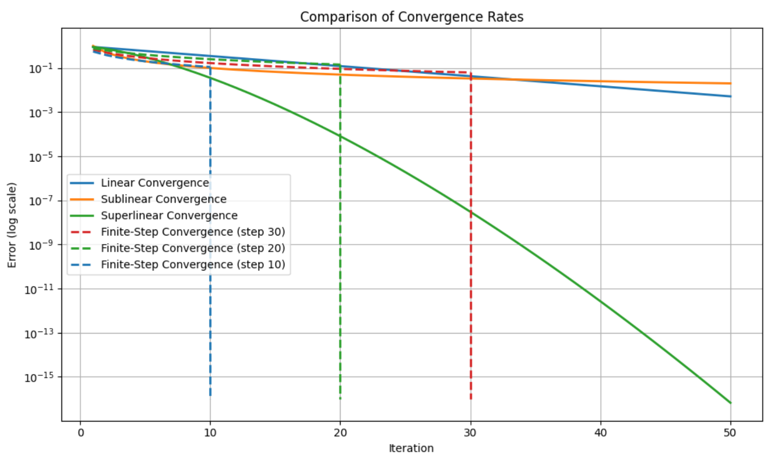

Figure F2. Comparison of convergence rates for optimization algorithms. The figure illustrates typical error decay behaviors: linear convergence (exponential decay), sublinear convergence (e.g., $1/k$), and superlinear convergence (faster than exponential). Additionally, three finite-step convergence scenarios are shown, where error drops to zero suddenly after a fixed number of iterations (at step 10, 20, and 30 respectively), representing idealized convergence in a finite number of steps. The y-axis is shown on a logarithmic scale to highlight differences in decay speed.

Reference:

[1] Nocedal, Jorge, and Stephen J. Wright, eds. Numerical optimization. New York, NY: Springer New York, 1999.

[2] Wright S J. Primal-dual interior-point methods[M]. Society for Industrial and Applied Mathematics, 1997.

[3] Trefethen L N, Bau D. Numerical linear algebra[M]. Society for Industrial and Applied Mathematics, 2022.

[4] Dennis Jr J E, Schnabel R B. Numerical methods for unconstrained optimization and nonlinear equations[M]. Society for Industrial and Applied Mathematics, 1996.

[5] Han, Ningning, Jian Lu, and Shidong Li. "The finite steps of convergence of the fast thresholding algorithms with f-feedbacks in compressed sensing." Numerical Algorithms 90.3 (2022): 1197-1223.