For find x_* s.t. $0 \in \partial f(x_*)$

we can solve $\begin{cases} -\frac{dx}{dt} \in \partial f(x) \\ x(0) \end{cases}$

① Explicit scheme (Subgradient Method)

$$x_{k+1} \in x_k - \eta \partial f(x_k)$$

$$x_{k+1} = x_k - \eta g_k, \quad g_k \in \partial f(x_k)$$

$$x_{k+1} = \arg\min_u f(x_k) + \langle g_k, u - x_k \rangle + \frac{1}{2\eta}\|u - x_k\|^2$$

$$\Leftrightarrow g_k + \frac{1}{\eta}(u_* - x_k) = 0$$

$$\Leftrightarrow u_* = x_k - \eta g_k$$

Newtons Method for smooth $f(x)$

$$x_{k+1} = x_k - \eta \underline{[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)} \qquad \nabla^2 f > 0$$

$$x_{k+1} = \arg\min_u f(x_k) + \langle \nabla f(x_k), u - x_k \rangle + \frac{1}{2\eta}(u - x_k)^T \nabla^2 f(x_k)(u - x_k)$$

$$\Leftrightarrow \nabla f(x_k) + \frac{1}{\eta}\nabla^2 f(x_k)(u_* - x_k) = 0$$

$$\Leftrightarrow u_* = x_k - \eta [\nabla f^2(x_k)]^{-1} \nabla f(x_k)$$

② Implicit Scheme (Proximal Point Method)

$$x_{k+1} \in x_k - \eta \partial f(x_{k+1})$$

$$x_{k+1} = (I + \eta \partial f)^{-1}(x_k)$$

$$x_{k+1} = \underset{u}{\operatorname{argmin}}\ f(u) + \frac{1}{2\eta}\|u - x_k\|^2$$

Theorem [Prox-Grad Inequality]

Assume
① $f(x)$ is convex
② $\nabla g(x)$ is L-continuous with L

$F(x) = f(x) + g(x)$

Let $\bar{y} = \mathrm{Prox}_f^{\eta}(y - \eta \nabla g(y))$, and $\eta \le \frac{1}{L}$

$$F(x) - F(\bar{y}) \ge \frac{L}{2}\|x - \bar{y}\|^2 - \frac{L}{2}\|x - y\|^2$$
$$+ g(x) - g(y) - \langle \nabla g(y), x - y \rangle$$

Corollary : Prox-Grad Inequality with $g(x) = 0$ implies

$$f(x) - f[\mathrm{Prox}_f^{\eta}(y)] \ge \frac{1}{2\eta}\|x - \mathrm{Prox}_f^{\eta}(y)\|^2 - \frac{1}{2\eta}\|x - y\|^2$$

Convergence of Proximal Point Method for convex $f(x)$:

① Prox-Grad Inequality $\Rightarrow$

(*) $$f(x_*) - f(x_{k+1}) \ge \frac{1}{2\eta}\|x_* - x_{k+1}\|^2 - \frac{1}{2\eta}\|x_* - x_k\|^2$$

$$\Rightarrow \|x_{k+1} - x_*\|^2 \leq \|x_k - x_*\|^2 \quad \forall \eta > 0$$

② Prox-Grad Inequality $\Rightarrow$

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{2\eta}\|x_k - x_{k+1}\|^2 \geq 0$$

$$\Rightarrow f(x_k) \searrow \qquad \forall \eta > 0$$

③ Sum $(*)$ $\boxed{f(x) - f[\mathrm{Prox}_f^{\eta}(y)] \geq \frac{1}{2\eta}\|x - \mathrm{Prox}_f^{\eta}(y)\|^2 - \frac{1}{2\eta}\|x - y\|^2}$

$$\sum_{k=0}^{n-1}[f(x_*) - f(x_{k+1})] \geq \frac{1}{2\eta}\|x_* - x_n\|^2 - \frac{1}{2\eta}\|x_* - x_0\|^2$$

$$n[f(x_n) - f(x_*)] \leq \sum_{k=0}^{n-1}[f(x_{k+1}) - f(x_*)] \leq \frac{1}{2\eta}\|x_* - x_0\|^2$$

$$\Rightarrow f(x_n) - f(x_*) \leq \frac{1}{2\eta}\|x_* - x_0\|^2 \cdot \frac{1}{n}$$

$$O\left(\frac{1}{k}\right), \quad \forall \eta > 0$$

④ Now assume $f(x)$ is strongly convex with μ.

The linear rate of Proximal-Gradient does NOT imply the same for proximal point.

Theorem $f(x)$ is strongly convex with $\mu > 0$

$$\Rightarrow f(x) \geq f(y) + \langle g, x - y\rangle + \frac{\mu}{2}\|x - y\|^2, \quad g \in \partial f(y)$$

$\boxed{\text{Theorem}}$ $f(x)$ is strongly convex with $\mu > 0$

$$\Rightarrow \quad \langle \partial f(x) - \partial f(y), x-y \rangle \geq \mu \|x-y\|^2$$

meaning:

$$\langle g-h, x-y \rangle \geq \mu \|x-y\|^2, \ \forall g \in \partial f(x), \ \forall h \in \partial f(y)$$

Proof: $f(x) \geq f(y) + \langle h, x-y \rangle + \frac{\mu}{2}\|x-y\|^2, h \in \partial f(y)$

$$f(y) \geq f(x) + \langle g, y-x \rangle + \frac{\mu}{2}\|y-x\|^2, g \in \partial f(x)$$

$$\Rightarrow \quad \langle g-h, x-y \rangle \geq \mu \|x-y\|^2$$

$\boxed{\text{Theorem}}$ $f(x)$ is strongly convex with $\mu > 0$

$$x_{k+1} = (I + \eta \partial f)^{-1}(x_k) \quad \forall \eta > 0$$

satiesfies

① $\|x_{k+1} - x_*\|^2 \leq \frac{1}{1+2\eta\mu}\|x_k - x_*\|^2$

② $\|x_k - x_*\|^2 \leq \left[\frac{1}{1+2\eta\mu}\right]^k \|x_0 - x_*\|^2$

② $f(x_k) - f(x_*) < \frac{1+2\eta\mu}{2\eta}\left[\frac{1}{1+2\eta\mu}\right]^k \|x_0 - x_*\|^2$

Proof: $x_k = x_{k+1} + \eta \cdot g_{k+1}, \ \exists g_{k+1} \in \partial f(x_{k+1})$

$$x_k - x_* = x_{k+1} - x_* + \eta \cdot g_{k+1}$$

$$\|x_k - x_*\|^2 = \|x_{k+1} - x_* + \eta \cdot g_{k+1}\|^2$$

$$= \|x_{k+1} - x_*\|^2 + \|\eta \cdot g_{k+1}\|^2$$

$$+ 2 \langle x_{k+1} - x_*, \eta\, g_{k+1} \rangle$$

$$\geq \|x_{k+1} - x_*\|^2 + 2 \langle x_{k+1} - x_*, \eta\, g_{k+1} \rangle$$

$$= \|x_{k+1} - x_*\|^2 + 2\eta \langle x_{k+1} - x_*, \partial f(x_{k+1}) - \partial f(x_*) \rangle$$

$$\geq \|x_{k+1} - x_*\|^2 + 2\eta\mu \|x_{k+1} - x_*\|^2$$

$$\Rightarrow \quad \|x_{k+1} - x_*\|^2 \leq \frac{1}{1+2\eta\mu} \|x_k - x_*\|^2$$

$$\Rightarrow \quad \|x_k - x_*\|^2 \leq \left[\frac{1}{1+2\eta\mu}\right]^k \|x_0 - x_*\|^2$$

Prox-Grad Inequality $\Rightarrow$

$$f(x_*) - f(x_{k+1}) \geq \frac{1}{2\eta} \|x_* - x_{k+1}\|^2 - \frac{1}{2\eta} \|x_* - x_k\|^2$$

$$\Rightarrow f(x_{k+1}) - f(x_*) \leq \frac{1}{2\eta} \|x_k - x_*\|^2$$

$$\leq \frac{1+2\eta\mu}{2\eta} \left[\frac{1}{1+2\eta\mu}\right]^{k+1} \|x_0 - x_*\|^2$$

Convergence Rate for nonsmooth problems

	Convexity	Strong Convexity
① Subgradient Method	$O(\frac{1}{\sqrt{k}})$	$O(\frac{1}{k})$
② Proximal Point Method	$O(\frac{1}{k})$	$O((\frac{1}{1+2\eta\mu})^k)$
③ Proximal Gradient	$O(\frac{1}{k})$	$O((1-\frac{\mu}{L})^k)$ $\eta=\frac{1}{L}$
④ Accelerated Prox Grad	$O(\frac{1}{k^2})$	$O((1-\sqrt{\frac{\mu}{L}})^k)$

Remark: ① Proximal Point is stable for any $\eta>0$.

② Proximal Point converges faster with larger $\eta>0$ for strongly convex $f(x)$.

Example: $f(x) = \frac{1}{2}x^T K x - x^T b$ $\quad K>0$

$K X_* = b$

$(I+\eta K)X_* = X_* + \eta b$

$X_* = (I+\eta K)^{-1}(X_* + \eta b)$

$$X_{k+1} = X_k - \eta \nabla f(X_{k+1})$$

$$= X_k - \eta [K X_{k+1} - b]$$

$$[I+\eta K] X_{k+1} = X_k + \eta b$$

$$X_{k+1} = [I+\eta K]^{-1}[X_k + \eta b]$$

$$X_{k+1} - X_* = [I+\eta K]^{-1}[X_k+\eta b] - X_*$$

$$= [I+\eta K]^{-1}[X_k - X_*]$$

$$e_{k+1} = [I+\eta K]^{-1} e_k \qquad \frac{1}{\eta}I + K$$

$$\|e_{k+1}\| \leq \|[I+\eta K]^{-1}\| \cdot \|e_k\|$$

$$\leq \frac{1}{1+\eta\lambda_1}\|e_k\|$$

eigenvalues of K: $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

Fast/Accelerated Proximal Gradient Method

$$\min_x f(x) + g(x) \qquad \begin{cases} \nabla g(x) \text{ is L-cont.} \\ f(x) \text{ is convex} \\ g(x) \text{ is convex} \end{cases}$$

Nesterov's Acceleration for $\min_x f(x)$ ∇f is L-cont.

$$\boxed{\begin{cases} \mathbf{x}_{k+1} &= \mathbf{y}_k - \frac{1}{L}\nabla f(\mathbf{y}_k) \\ \mathbf{y}_{k+1} &= \mathbf{x}_{k+1} + \frac{t_k - 1}{t_{k+1}}(\mathbf{x}_{k+1} - \mathbf{x}_k) \end{cases} \qquad \mathbf{x}_0 = \mathbf{y}_0,}$$

$t_k^2 - t_k \leq t_{k-1}$; Example: $t_k = \frac{k+1}{2}$

$$\begin{cases} x_{k+1} = (I + \frac{1}{L}\partial f)^{-1}(y_k - \frac{1}{L}\nabla g(y_k)) \\ t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2} \Rightarrow t_{k+1}^2 - t_k = t_k \qquad x_0 = y_0 \\ y_{k+1} = x_{k+1} + \frac{t_k - 1}{t_{k+1}}(x_{k+1} - x_k) \end{cases}$$

$$\min_{x} \|x\|_1 + \lambda\|Ax-b\|^2$$