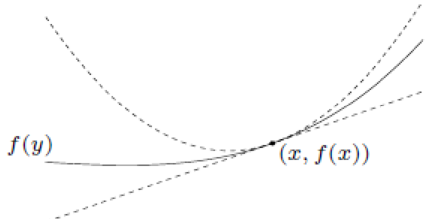
First order methods for composite functions

Prox method with nonsmooth term

- Consider: $\boxed{\min_x F(x) = f(x) + g(x)}$

$$|\nabla f(x) - \nabla f(y)| \leq L||x - y||$$

- Quadratic upper approximation



$$f(y)+g(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2\mu}||y - x||^2+g(y) = Q_{f,\mu}(x, y)$$

$$F(y) \leq f(x) + \frac{1}{2\mu}||x - \mu\nabla f(x)^\top - y||^2 - \frac{1}{\mu}\|\nabla f(x^k)\|^2+g(y)$$

Assume that *g(y)* is such that the above function is easy to optimize over *y*

Example 1 Sparse optimization

$$\boxed{\min_x f(x) + ||x||_1}$$

- Minimize upper approximation function $Q_{f,\mu}(x,y)$ on each iteration

$$\min_y Q_{f,\mu}(x,y) = \min_y f(x) + \frac{1}{2\mu}||x - \mu\nabla f(x)^\top - y||^2 + ||y||_1$$

$$\Updownarrow$$

Shrinkage operator

$$\sum_i \min_{y_i} \left[\frac{1}{2\mu}(y_i - r_i)^2 + |y_i|\right]$$

$$\Updownarrow$$

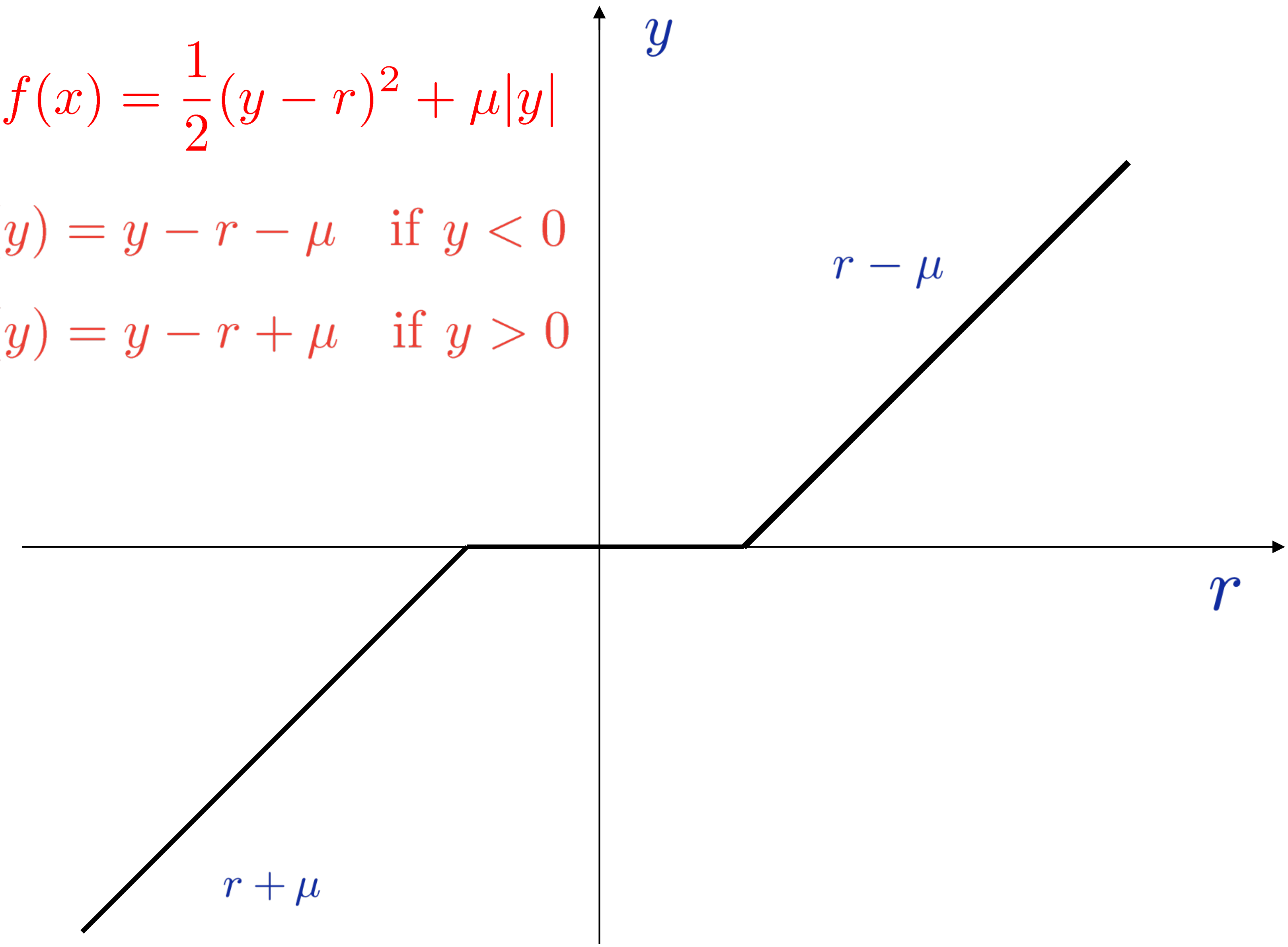
Closed form solution! $O(n)$ effort

$$\min_{y_i} \frac{1}{2}(y_i - r_i)^2 + \mu|y_i| \rightarrow y_i^* = \begin{cases} r_i - \mu & \text{if } r_i > \mu \\ 0 & \text{if } -\mu \leq r_i \leq \mu \\ r_i + \mu & \text{if } r_i < -\mu \end{cases}$$

$$f(x) = \frac{1}{2}(y - r)^2 + \mu|y|$$

$$f'(y) = y - r - \mu \quad \text{if } y < 0$$

$$f'(y) = y - r + \mu \quad \text{if } y > 0$$



ISTA/Gradient prox method

$$\boxed{\min_x F(x) = f(x) + g(x)}$$

- Minimize quadratic upper approximation on each iteration

$$x^{k+1} = \mathrm{argmin}_y Q_f(x^k, y)$$

$$Q_{f,\mu}(x, y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2\mu}||y - x||^2 + g(y)$$

- If $\mu \leq 1/L$ then in $O(L/\epsilon)$ iterations finds solution

$$\bar{x} : \; F(\bar{x}) \leq F(x^*) + \epsilon$$

Fast first-order method

Nesterov, Beck & Teboulle

$$\boxed{\min_x F(x) = f(x) + g(x)}$$

- Minimize upper approximation at an "accelerated" point.

$$x^k = \text{argmin}_y Q_f(y^k, y)$$

$$t_{k+1} := (1 + \sqrt{1 + 4t_k^2})/2$$
$$y^{k+1} := x^k + \tfrac{t_k - 1}{t_{k+1}}[x^k - x^{k-1}]$$

- If $\mu \leq$ *1/L* then in $O(\sqrt{L/\epsilon})$ iterations finds solution

$$\bar{x} : \; F(\bar{x}) \leq F(x^*) + \epsilon$$

Practical first order algorithms using backtracking search

Iterative Shrinkage Threshholding Algorithm (ISTA)

$$\boxed{\min_x F(x) = f(x) + g(x)}$$

- Minimize quadratic upper relaxation on each iteration

$$x^{k+1} = \operatorname{argmin}_y f(x^k) + \frac{1}{2\mu_k}||x^k - \mu_k \nabla f(x^k)^\top - y||^2 + g(y)$$

- Using line search find μ_k such that

$$F(x^{k+1}) \leq Q_f(x^k, x^{k+1})$$

- In *$O(1/\mu_{min}\epsilon)$* iterations finds ϵ-optimal solution (in practice better)

Nesterov, 07
Beck&Teboulle, Tseng,
Auslender&Teboulle, 08

Fast Iterative Shrinkage Threshholding Algorithm (FISTA)

$$\boxed{\min_x F(x) = f(x) + g(x)}$$

- Minimize quadratic upper relaxation on each iteration

$$x^k = \operatorname{argmin}_y Q_f(y^k, y) = f(y^k) + \frac{1}{2\mu_k}||y^k - \mu_k \nabla f(y^k)^\top - y||^2 + g(y)$$

- Using line search find $\boxed{\mu_k \leq \mu_{k-1}}$ such that

Can be restrictive

$$F(x^k) \leq Q_f(y^k, x^k)$$

$$t_{k+1} := \quad (1 + \sqrt{1 + 4t_k^2})/2$$

$$y^{k+1} := \quad x^k + \tfrac{t_k - 1}{t_{k+1}}[x^k - x^{k-1}]$$

- In $O(\sqrt{1/\mu_{min}\epsilon})$ iterations finds ϵ-optimal solution

Nesterov, Beck&Teboulle, Tseng