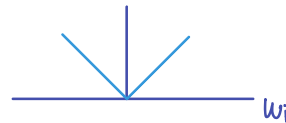
Proximal Gradient Descent Algorithm

Proximal Gradient Descent Solves regularized least-squares problems.

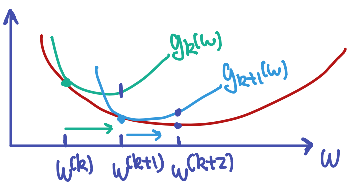
$\min_w \|Aw - d\|_2^2 + \lambda r(w)$ $r(w)$: regularizer, $\lambda > 0$ tuning parameter

Example of Convex Regularizers

- Ridge (Tikhonov) $r(w) = \|w\|_2^2 = \sum_{i=1}^{M} w_i^2$
- LASSO (ℓ_1) $r(w) = \|w\|_1 = \sum_{i=1}^{M} |w_i|$ not differentiable



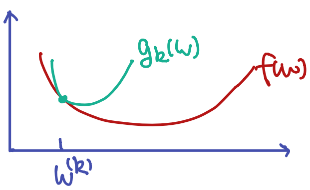
Proximal Gradient Descent Concept



$f(w) = \|Aw - d\|_2^2 + \lambda r(w)$

- solve sequence of simpler problems
- simple for separable $r(w) = \sum_i h_i(w_i)$

Find $g_k(w)$ so $f(w) \le g_k(w)$, $g_k(w^{(k)}) = f(w^{(k)})$



minimize $g_k(w) \Longrightarrow f(w)$ decreases

$$f(w) = \|d - Aw\|_2^2 + \lambda r(w)$$
$$= \|(d - Aw^{(k)}) + (Aw^{(k)} - Aw)\|_2^2 + \lambda r(w)$$
$$= \underbrace{\|d - Aw^{(k)}\|_2^2}_{C_k} + \underbrace{\|A(w^k - w)\|_2^2}_{\le \|A\|_{op}^2 \|w^k - w\|_2^2} + 2\underbrace{(d - Aw^{(k)})^T A}_{V_k^T}(w^{(k)} - w) + \lambda r(w)$$

Define step size: $0 < \tau < \frac{1}{\|A\|_{op}^2} \Rightarrow \frac{1}{\tau} > \|A\|_{op}^2$

$$f(w) \le g_k(w) = C_k + \frac{1}{\tau}\|w^k - w\|_2^2 + 2V_k^T(w^k - w) + \lambda r(w)$$

$g_k(w)$ is separable

for $r(w)$ separable: $g_k(w) = C_k + \sum_{i=1}^{M} q_i(w_i)$ no $w_i w_j$ terms

Find $w^{(k+1)} = \arg\min_w g_k(w)$

$$g_k(w) = C_k + \frac{1}{\tau}\|w^k - w\|_2^2 + 2V_k^T(w^k - w) + \lambda r(w)$$
$$\tau g_k(w) = \tau C_k + \|w^k - w\|_2^2 + 2\tau V_k^T(w^k - w) + \tau\lambda r(w)$$
$$= \tau C_k - \tau^2 V_k^T V_k + (\tau V_k + (w^k - w))^T(\tau V_k + (w^k - w)) + \tau\lambda r(w)$$

$z^{(k)}$

$$w^{(k+1)} = \arg\min_{w} \| z^{k} - w \|_2^2 + \lambda \tau r(w)$$

$$z^{(k)} = w^{(k)} + \tau V_k$$

$$= w^{(k)} + \tau A^T (d - A w^{(k)})$$

$$= w^{(k)} - \tau A^T (A w^{(k)} - d)$$

Least-squares gradient descent (Landweber)

Alternative LS gradient descent and regularization

$w^{(0)} = 0 \,,\; 0 < \tau < \frac{1}{\|A\|_{op}^2}$ initialize

$z^{(k)} = w^{(k)} - \tau A^T (Aw - d)$ LS Gradient Descent

$w^{(k+1)} = \arg\min_{w} \| z^{(k)} - w \|_2^2 + \lambda \tau r(w)$ Regularize

if $\| w^{k+1} - w \| < \varepsilon$, stop Check if converged

Regularization simple for $r(w)$ separable

if $r(w) = \sum_{i=1}^{M} h_i(w_i)$

$$w^{(k+1)} = \arg\min_{w_i,\, i=1\cdots M} \sum_{i=1}^{M} \left((z_i^{(k)} - w_i)^2 + \lambda \tau h(w_i) \right)$$

. M scalar minimizations