Regularization and Stability

Mahbub Majumdar

BRAC University
66 Mohakhali
Dhaka, Bangladesh

May 28, 2018

Table of Contents

To solve a practical problem:

- Calculate a physical quantity perturbatively in g.

$$A(g) = a_0 + a_1 g + a_2 g^2 + \cdots$$

- Renormalize the individual terms in some way, so each term is finite.
- Physicists use loops to index the terms.
- Hope that the sum converges.
- The a_i decrease for $i < N$, but grow for $i \geq N$.
- Interesting series never converge! Truncate the series at N.
- Treat the series as an asymptotic series.

$$\underset{\mathbf{w}}{\operatorname{argmin}} \left(L_S(\mathbf{w}) + R(\mathbf{w}) \right)$$

$$A(S) = \operatorname*{argmin}_{\mathbf{w}}(L_S(\mathbf{w}) + \lambda\|\mathbf{w}^2\|)$$

$$\underset{\mathbf{w}\in\mathbb{R}^d}{\operatorname{argmin}} \left(\lambda \|\mathbf{w}_2^2\| + \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} (\langle \mathbf{w}, \mathbf{x}_i \rangle - y_i)^2 \right)$$

$$(2\lambda m I + A)\mathbf{w} = \mathbf{b}$$

$$A = \left(\sum_{i=1}^{m} \mathbf{x}_i \mathbf{x}_i^T \right) \quad \text{and} \quad \mathbf{b} = \sum_{i=1}^{m} y_i \mathbf{x}_i$$

$$\mathbf{w} = (2\lambda m I + A)^{-1}\mathbf{b}$$

$$\mathop{\mathbb{E}}_{S \sim \mathcal{D}^m} [L_{\mathcal{D}}(A(S))] \leq \min_{\mathbf{w} \in \mathcal{H}} L_{\mathcal{D}}(\mathbf{w}) + \epsilon$$

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}(A(S)) - L_S(A(S))\right] \\
&= \mathop{\mathbb{E}}_{(S,z')\sim\mathcal{D}^{m+1},\, i\sim U(m)}\left[\ell(A(S^{(i)}, z_i)) - \ell(A(S), z_i)\right]
\end{aligned}
$$

$$
\mathop{\mathbb{E}}_{S}\left[L_{\mathcal{D}}(A(S))\right] = \mathop{\mathbb{E}}_{S,z'}\left[\ell(A(S), z')\right] = \mathop{\mathbb{E}}_{S,z'}\left[\ell(A(S^{(i)}), z_i)\right]
$$

$$
\mathop{\mathbb{E}}_{S}\left[L_S(A(S))\right] = \mathop{\mathbb{E}}_{S,i}\left[\ell\left(A(S), z_i\right)\right]
$$

$$\mathop{\mathbb{E}}_{(S,z')\sim\mathcal{D}^{m+1},i\sim U(m)}\left[\ell\left(A\left(S^{(i)},z_i\right)\right)-\ell\left(A(S),z_i\right)\right]\leq\epsilon(m)$$

$$f(\alpha \mathbf{w} + (1-\alpha)\mathbf{u}) \leq \alpha f(\mathbf{w}) + (1-\alpha) f(\mathbf{u}) - \frac{\lambda}{2}\alpha(1-\alpha)\|\mathbf{w}-\mathbf{u}\|^2$$

$$f(\mathbf{w}) - f(\mathbf{u}) \geq \frac{\lambda}{2}\|\mathbf{w} - \mathbf{u}\|^2$$

$$\frac{f(\mathbf{u} + \alpha(\mathbf{w} - \mathbf{u})) - f(\mathbf{u})}{\alpha} \leq f(\mathbf{w}) - f(\mathbf{u}) - \frac{\lambda}{2}(1 - \alpha)\|\mathbf{w} - \mathbf{u}\|^2$$

$$f_S(\mathbf{v}) - f_S(A(S)) \geq \lambda \|\mathbf{v} - A(S)\|^2$$

$$
\begin{aligned}
f_S(\mathbf{v}) - f_S(\mathbf{u}) &= L_S(\mathbf{v}) + \lambda\|\mathbf{v}\|^2 - \left(L_S(\mathbf{u}) + \lambda\|\mathbf{u}\|^2\right) \\
&= L_{S^{(i)}}(\mathbf{v}) + \lambda\|\mathbf{v}\|^2 - \left(L_{S^{(i)}}(\mathbf{u}) + \lambda\|\mathbf{u}\|^2\right) \\
&\quad + \frac{\ell\left(\mathbf{v}, z_i\right) - \ell\left(\mathbf{u}, z_i\right)}{m} + \frac{\ell\left(\mathbf{u}, z'\right) - \ell\left(\mathbf{v}, z'\right)}{m}
\end{aligned}
$$

$$f_S(A(S^{(i)})) - f_S(A(S))$$
$$\leq \frac{\ell\left(A\left(S^{(i)}\right), z_i\right) - \ell\left(A(S), z_i\right)}{m} + \frac{\ell\left(A(S), z'\right) - \ell\left(A\left(S^{(i)}\right), z'\right)}{m}$$

$$
\begin{aligned}
&\lambda \|A(S^{(i)}) - A(S)\|^2 \\
&\leq \frac{\ell\left(A\left(S^{(i)}\right), z_i\right) - \ell\left(A(S), z_i\right)}{m} + \frac{\ell\left(A(S), z'\right) - \ell\left(A\left(S^{(i)}\right), z'\right)}{m}
\end{aligned}
$$

$$\ell(A(S^{(i)}), z_i) - \ell(A(S), z_i) \leq \rho \|A(S^{(i)}) - A(S)\|$$

$$\ell(A(S), z') - \ell(A(S^{(i)}), z') \leq \rho \|A(S^{(i)}) - A(S)\|$$

$$\lambda \|A(S^{(i)}) - A(S)\|^2 \leq \frac{2\rho \|A(S^{(i)}) - A(S)\|}{m}$$

$$\|A(S^{(i)}) - A(S)\| \leq \frac{2\rho}{\lambda m}$$

$$\ell(A(S^{(i)}), z_i) - \ell(A(S), z_i) \leq \frac{2\rho^2}{\lambda m}$$

$$\mathop{\mathbb{E}}_{S\sim\mathcal{D}^m}\left[L_{\mathcal{D}}(A(S)) - L_S(A(S))\right] \leq \frac{2\rho^2}{\lambda m}$$

$$\|\nabla f(\mathbf{w})\|^2 \leq 2\beta f(\mathbf{w})$$

$$
\begin{aligned}
\ell(A(S^{(i)}), z_i) - \ell\left(A(S), z_i\right) \leq & \left\langle \nabla \ell\left(A(S), z_i\right), A(S^{(i)}) - A(S) \right\rangle \\
& + \frac{\beta}{2}\|A(S^{(i)}) - A(S)\|^2
\end{aligned}
$$

$$
\begin{aligned}
\ell(A(S^{(i)}), z_i) - \ell\left(A(S), z_i\right) \\
&\le \|\nabla \ell\left(A(S), z_i\right)\| \|A(S^{(i)}) - A(S)\| + \frac{\beta}{2}\|A(S^{(i)}) - A(S)\|^2 \\
&\le \sqrt{2\beta\ell(A(S), z_i)}\|A(S^{(i)}) - A(S)\| + \frac{\beta}{2}\|A(S^{(i)}) - A(S)\|^2
\end{aligned}
$$

$$
\begin{aligned}
\ell\left(A(S), z'\right) - \ell(A(S^{(i)}), z') \\
&\le \sqrt{2\beta\ell\left(A(S^{(i)}), z'\right)}\|A(S^{(i)}) - A(S)\| + \frac{\beta}{2}\|A(S^{(i)}) - A(S)\|^2
\end{aligned}
$$

$$\|A(S^{(i)}) - A(S)\| \leq \frac{\sqrt{2\beta}}{(\lambda m - \beta)} \left(\sqrt{\ell\left(A(S), z_i\right)} + \sqrt{\ell\left(A\left(S^{(i)}\right), z'\right)} \right)$$

$$\|A(S^{(i)}) - A(S)\| \leq \frac{\sqrt{8\beta}}{\lambda m} \left(\sqrt{\ell\left(A(S), z_i\right)} + \sqrt{\ell\left(A\left(S^{(i)}\right), z'\right)} \right)$$

$$
\begin{aligned}
&\ell(A(S^{(i)}), z_i) - \ell\left(A(S), z_i\right) \\
&\leq \sqrt{2\beta\ell\left(A(S^{(i)}), z'\right)}\|A\left(S^{(i)}\right) - A(S)\| + \frac{\beta}{2}\|A\left(S^{(i)}\right) - A(S)\|^2 \\
&\leq \left(\frac{4\beta}{\lambda m} + \frac{8\beta^2}{(\lambda m)^2}\right)\left(\sqrt{\ell(A(S), z')} + \sqrt{\ell(A(S^{(i)}), z')}\right)^2 \\
&\leq \frac{8\beta}{\lambda m}\left(\sqrt{\ell(A(S), z')} + \sqrt{\ell(A(S^{(i)}), z')}\right)^2 \\
&\leq \frac{24\beta}{\lambda m}\left(\ell(A(S), z') + \ell(A(S^{(i)}), z')\right)
\end{aligned}
$$

$$\mathbb{E}\left[\ell\left(A\left(S^{(i)}\right), z_i\right) - \ell\left(A(S), z_i\right)\right] \leq \frac{48\beta}{\lambda m}\mathbb{E}\left[L_S(A(S))\right]$$

$$L_S(A(S)) \leq L_S(A(S)) + \lambda\|A(S)\|^2 \leq L_S(\mathbf{0}) + \lambda\|\mathbf{0}\|^2 = L_S(\mathbf{0}) \leq C$$

$$\mathbb{E}\left[\ell(A(S^{(i)}), z_i) - \ell(A(S), z_i)\right] \leq \frac{48\beta C}{\lambda m}$$

$$\mathop{\mathbb{E}}_{S}[L_D(A(S))] = \mathop{\mathbb{E}}_{S}[L_S(A(S))] + \mathop{\mathbb{E}}_{S}[L_D(A(S)) - L_S(A(S))]$$

$$L_S(A(S)) \leq L_S(A(S)) + \lambda\|A(S)\|^2 \leq L_S(\mathbf{w}^*) + \lambda\|\mathbf{w}^*\|^2$$

$$\mathop{\mathbb{E}}_{S}\left[L_S(A(S))\right] \leq L_{\mathcal{D}}\left(\mathbf{w}^*\right) + \lambda\|\mathbf{w}^*\|^2$$

$$\mathop{\mathbb{E}}_{S}\left[L_D(A(S))\right] \leq L_{\mathcal{D}}\left(\mathbf{w}^*\right) + \lambda\|\mathbf{w}^*\|^2 + \mathop{\mathbb{E}}_{S}\left[L_D(A(S)) - L_S(A(S))\right]$$

$$\forall \mathbf{w}^*, \underset{S}{\mathbb{E}}\left[L_S(A(S))\right] \leq L_{\mathcal{D}}\left(\mathbf{w}^*\right) + \lambda \|\mathbf{w}^*\|^2 + \frac{2\rho^2}{\lambda m}$$

$$\mathbb{E}_{S}[L_{\mathcal{D}}(A(S))] \leq \min_{w \in \mathcal{H}} L_{\mathcal{D}}(\mathbf{w}) + \rho B \sqrt{\frac{8}{m}}$$

$$\mathop{\mathbb{E}}_{S}[L_{\mathcal{D}}(A(S))] \leq \left(1 + \frac{48\beta}{\lambda m}\right) \mathop{\mathbb{E}}_{S}[L_S(A(S))] \leq \left(1 + \frac{48\beta}{\lambda m}\right) \left(L_{\mathcal{D}}\left(w^*\right) + \lambda \|\mathbf{w}^*\right.$$

$$\mathop{\mathbb{E}}_{S}\left[L_{\mathcal{D}}(A(S))\right] \leq \min_{w \in \mathcal{H}} L_{\mathcal{D}}(w) + \epsilon$$