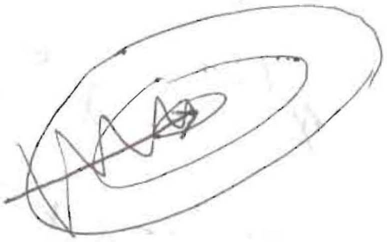
Nonlinear Optimization in Machine Learning

Lecture 3 Gradient Methods using Momentum

The steepest descent method is "greedy" in that it steps in the direction that is most productive at current iterate — no explicit use of knowledge of f at earlier iterations

adding "momentum" — search direction tends to be similar to that one used in the previous step



- the heavy ball method
- the conjugate gradient method
- Nesterov's accelerated gradient method

Physical Intuition $$\frac{dx}{dt} = -\nabla f(x)$$

$$\mu \frac{d^2x}{dt^2} = -\nabla f(x) - \mu b \frac{dx}{dt}$$

"Small mass limit" $\mu \to 0$ "regularization"

Finite difference approximation

$$\mu \frac{x(t+\Delta t) - 2x(t) + x(t-\Delta t)}{(\Delta t)^2} \approx -\nabla f(x(t)) - \frac{\mu b\,(x(t+\Delta t) - x(t))}{\Delta t}$$

i.e.

$$x(t+\Delta t) = x(t) - \alpha \nabla f(x(t)) + \beta(x(t) - x(t-\Delta t))$$

"Heavy ball method" (Polyak 1964)

$$x^{k+1} = x^k - \alpha \nabla f(x^k) + \beta(x^k - x^{k-1})$$

$$x^{-1} := x^0$$

or

$$x^{k+1} = x^k + p^k$$

$$p^k = -\alpha \nabla f(x^k) + \beta p^{k-1}$$

$$p^{-1} = 0$$

"Nesterov's accelerated gradient method"

$$x^{k+1} = x^k - \alpha \nabla f(x^k + \beta(x^k - x^{k-1})) + \beta(x^k - x^{k-1})$$

$$x^{-1} := x^0$$

or
$$y^k = x^k + \beta_k (x^k - x^{k-1}) \qquad y^0 = x^0$$
$$x^{k+1} = y^k - \alpha_k \nabla f(y^k)$$

Convergence of Nesterov's method for Convex Quadratics

$$f(x) = \frac{1}{2} x^T Q x - b^T x + c$$

$$Q \succ 0 \quad , \quad 0 < m = \lambda_n \le \lambda_{n-1} \le \dots \le \lambda_2 \le \lambda_1 = L$$

Condition Number $\kappa = \frac{L}{m}$

Minimiser of f is $x^* = Q^{-1} b$

$$\nabla f(x) = Qx - b = Q(x - x^*)$$

"Spectral method"

$$x^{k+1} - x^* = (x^k - x^*) - \alpha Q(x^k + \beta(x^k - x^{k-1}) - x^*) + \beta\left((x^k - x^*) - (x^{k-1} - x^*)\right)$$

$$\underbrace{\begin{bmatrix} x^{k+1} - x^* \\ x^k - x^* \end{bmatrix}}_{w^k} = \underbrace{\begin{bmatrix} (1+\beta)(I-\alpha Q) & -\beta(I-\alpha Q) \\ I & 0 \end{bmatrix}}_{T} \underbrace{\begin{bmatrix} x^k - x^* \\ x^{k-1} - x^* \end{bmatrix}}_{w^{k-1}}$$

$$w^k = T w^{k-1} \qquad k = 1, 2, \ldots$$

$$w^0 = \begin{bmatrix} x^0 - x^* \\ x^0 - x^* \end{bmatrix}$$

How to calculate the eigenvalues of T?

<u>Theorem</u> $\alpha = \frac{1}{L}$ $\quad \beta = \frac{\sqrt{L} - \sqrt{m}}{\sqrt{L} + \sqrt{m}} = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}$

eigenvalues of T are

$$\nu_{i,1} = \frac{1}{2}\left[(1+\beta)(1-\alpha\lambda_i) + i\sqrt{4\beta(1-\alpha\lambda_i) - (1+\beta)^2(1-\alpha\lambda_i)^2}\right]$$

$$\nu_{i,2} = \frac{1}{2}\left[(1+\beta)(1-\alpha\lambda_i) - i\sqrt{4\beta(1-\alpha\lambda_i) - (1+\beta)^2(1-\alpha\lambda_i)^2}\right]$$

$$\rho(T) \leq 1 - \frac{1}{\sqrt{\kappa}}$$

Proof. $Q = U \Lambda U^T \quad \Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$

$$\begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}^T \begin{bmatrix} (1+\beta)(I-\alpha Q) & -\beta(I-\alpha Q) \\ I & 0 \end{bmatrix} \begin{bmatrix} U & 0 \\ 0 & U \end{bmatrix}$$

$$= \begin{bmatrix} (1+\beta)(I-\alpha\Lambda) & -\beta(I-\alpha\Lambda) \\ I & 0 \end{bmatrix}$$

"permute" $\rightsquigarrow \begin{bmatrix} T_1 & & 0 \\ & T_2 & \\ 0 & & \ddots \; T_n \end{bmatrix}$

$$T_i = \begin{bmatrix} (1+\beta)(1-\alpha\lambda_i) & -\beta(1-\alpha\lambda_i) \\ 1 & 0 \end{bmatrix} \quad i = 1, 2, \ldots, n$$

$$\det(-\nu I + T_i) = [(1+\beta)(1-\alpha\lambda_i) - \nu][-\nu] + \beta(1-\alpha\lambda_i)$$

$$= \nu^2 - \nu(1+\beta)(1-\alpha\lambda_i) + \beta(1-\alpha\lambda_i)$$

So $\nu = \frac{1}{2}\left[(1+\beta)(1-\alpha\lambda_i) \pm i\sqrt{4\beta(1-\alpha\lambda_i) - (1+\beta)^2(1-\alpha\lambda_i)^2}\right]$

$\lambda_i \in (m, L)$

$$|\nu| = \frac{1}{2}\sqrt{(1+\beta)^2(1-\alpha\lambda_i)^2 + 4\beta(1-\alpha\lambda_i) - (1+\beta)^2(1-\alpha\lambda_i)^2}$$

$$= \frac{1}{2}\sqrt{4\beta(1-\alpha\lambda_i)} = \sqrt{\beta}\sqrt{1 - \frac{\lambda_i}{L}} \quad \text{when } \alpha = \frac{1}{L}$$

$$\sqrt{\beta}\sqrt{1-\frac{\lambda_i}{L}} \le \sqrt{\beta}\sqrt{1-\frac{m}{L}} = \left(\frac{\sqrt{L}-\sqrt{m}}{\sqrt{L}+\sqrt{m}}\ \frac{L-m}{L}\right)^{1/2}$$

$$= \frac{\sqrt{L}-\sqrt{m}}{\sqrt{L}} = 1-\sqrt{\frac{m}{L}} = 1-\frac{1}{\sqrt{\kappa}}$$

#

$$\rho(T) = \lim_{k\to\infty} \left(\|T^k\|\right)^{1/k}$$

$$\|w^k\| \le C\|w^0\|\left(\rho(T)+\varepsilon\right)^k \quad \text{for } \forall \varepsilon > 0$$

Comparison of the rates with steepest descent method

$\alpha = \frac{1}{L}$ GD Needs $O\left(\frac{L}{m}|\log\varepsilon|\right)$ to reach $f(x^k)-f^* \le \varepsilon$

Nesterov Needs $O\left(\sqrt{\frac{L}{m}}|\log\varepsilon|\right)$ to reach a factor ε in $\|w^k\|$

depends on how large κ is!

"Lyapunov method"

Lyapunov function $V: \mathbb{R}^d \to \mathbb{R}$

$V(w) > 0$ for all $w \neq w^* \in \mathbb{R}^d$

$V(w^*) = 0$

$$V(w^{k+1}) < \rho^2 V(w^k) \qquad 0<\rho<1$$

$$V(w) = w^T P w \qquad P \succ 0$$

Want $(w^k)^T P w^k < \rho^2 (w^{k-1})^T P w^{k-1}$ $\quad (*)$

Assuming $(*)$ we obtain

$$(w^k)^T P w^k < \rho^{2k} (w^0)^T P w^0$$

i.e. $\lambda_{min}(P) \|x^k - x^*\|^2 \leq \lambda_{min}(P) \|w^k\|^2$

$$\leq \rho^{2k} \|P\| \cdot \|w^0\|^2$$

$$= 2\rho^{2k} \|P\| \cdot \|x^0 - x^*\|^2$$

so $\quad \|x^k - x^*\| \leq \sqrt{2\,\mathrm{cond}(P)}\, \|x^0 - x^*\| \rho^k$

How to guarentee $(*)$? Want $\rho(T) < \rho$

$$\rho(T) < \rho \iff \exists P \succ 0 \text{ s.t. } T^T P T - \rho^2 P \prec 0$$

What happens if we want a Lyapunov function at the non-quadratic case?

Convergence for strongly convex functions

$f(x) - f^*$ is a Lyapunov function for the steepest descent method

$$\tilde{x}^k = x^k - x^*, \quad \tilde{y}^k = y^k - y^* \quad \text{etc.}$$

$$V_k = f(x^k) - f^* + \frac{L}{2} \| \tilde{x}^k - \rho^2 \tilde{x}^{k-1} \|^2$$

Aim to show $V_{k+1} \le \rho^2 V_k$ for some $0 < \rho < 1$

$$\alpha_k = \frac{1}{L} \qquad \beta_k = \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}$$

$$f(x^{k+1}) \le f(y^k) - \frac{L}{2} \left\| \frac{1}{L} \nabla f(y^k) \right\|^2$$

actually

$$f(x^{k+1}) \le f(y^k) - \nabla f(y^k)^T (y^k - x^{k+1}) + \frac{L}{2} \| y^k - x^{k+1} \|^2$$

$$= f(y^k) - \frac{1}{L} \| \nabla f(y^k) \|^2 + \frac{1}{2L} \| \nabla f(y^k) \|^2$$

$$= f(y^k) - \frac{1}{2L} \| \nabla f(y^k) \|^2$$

(Nesterov
$$\begin{cases} y^k = x^k + \dfrac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} (x^k - x^{k-1}) \\ x^{k+1} = y^k - \dfrac{1}{L} \nabla f(y^k) \end{cases} \qquad y^0 = x^0$$
)

$$u^k = \frac{1}{L}\nabla f(y^k) \qquad u^* = 0 \quad , \quad \tilde{u}^k = u^k$$

$$V_{k+1} = f(x^{k+1}) - f^* + \frac{L}{2}\|\tilde{x}^{k+1} - \rho^2\tilde{x}^k\|^2$$

$$\leq f(y^k) - f^* - \frac{L}{2}\|\tilde{u}^k\|^2 + \frac{L}{2}\|\tilde{x}^{k+1} - \rho^2\tilde{x}^k\|^2$$

$$= \rho^2\left[f(y^k) - f^* + L(\tilde{u}^k)^T(\tilde{x}^k - \tilde{y}^k)\right]$$

$$- \rho^2 L(\tilde{u}^k)^T(\tilde{x}^k - \tilde{y}^k) + (1-\rho^2)\left(f(y^k) - f^* - L(\tilde{u}^k)^T\tilde{y}^k\right)$$

$$+ (1-\rho^2)L(\tilde{u}^k)^T\tilde{y}^k - \frac{L}{2}\|\tilde{u}^k\|^2 + \frac{L}{2}\|\tilde{x}^{k+1} - \rho^2\tilde{x}^k\|^2$$

We have

$$f(y^k) \leq f(x^k) - \nabla f(y^k)^T(x^k - y^k) - \frac{m}{2}\|x^k - y^k\|^2$$

$$= f(x^k) - L(\tilde{u}^k)^T(\tilde{x}^k - \tilde{y}^k) - \frac{m}{2}\|\tilde{x}^k - \tilde{y}^k\|^2$$

$$f(x^*) \geq f(y^k) + \nabla f(y^k)^T(x^* - y^k) + \frac{m}{2}\|y^k - x^*\|^2$$

$$= f(y^k) - L(\tilde{u}^k)^T\tilde{y}^k + \frac{m}{2}\|\tilde{y}^k\|^2$$

So

$$V_{k+1} \le \rho^2\left(f(x^k) - f^* - \frac{m}{2}\|\tilde{x}^k - \tilde{y}^k\|^2\right) - \frac{m(1-\rho^2)}{2}\|\tilde{y}^k\|^2$$

$$- \rho^2 L(\tilde{u}^k)^T(\tilde{x}^k - \tilde{y}^k) + (1-\rho^2) L(\tilde{u}^k)^T \tilde{y}^k$$

$$- \frac{L}{2}\|\tilde{u}^k\|^2 + \frac{L}{2}\|\tilde{x}^{k+1} - \rho^2 \tilde{x}^k\|^2$$

$$= \rho^2\left[f(x^k) - f^* + \frac{L}{2}\|\tilde{x}^k - \rho^2 \tilde{x}^{k-1}\|^2\right]$$

$$- \frac{m\rho^2}{2}\|\tilde{x}^k - \tilde{y}^k\|^2 - \frac{m(1-\rho^2)}{2}\|\tilde{y}^k\|^2 + L(\tilde{u}^k)^T(\tilde{y}^k - \rho^2 \tilde{x}^k)$$

$$- \frac{L}{2}\|\tilde{u}^k\|^2 + \frac{L}{2}\|\tilde{x}^{k+1} - \rho^2 \tilde{x}^k\|^2 - \frac{\rho^2 L}{2}\|\tilde{x}^k - \rho^2 \tilde{x}^{k-1}\|^2$$

$$= \rho^2 V_k + R_k$$

where

$$R_k = -\frac{m\rho^2}{2}\|\tilde{x}^k - \tilde{y}^k\|^2 - \frac{m(1-\rho^2)}{2}\|\tilde{y}^k\|^2 + L(\tilde{u}^k)^T(y^k - \rho^2 \tilde{x}^k)$$

$$- \frac{L}{2}\|\tilde{u}^k\|^2 + \frac{L}{2}\|\tilde{x}^{k+1} - \rho^2 \tilde{x}^k\| - \frac{\rho^2 L}{2}\|\tilde{x}^k - \rho^2 \tilde{x}^{k-1}\|^2$$

$$\rho^2 = 1 - \frac{1}{\sqrt{\kappa}}$$

$$R_k = -\frac{1}{2} L \rho^2 \left(\frac{1}{\kappa} + \frac{1}{\sqrt{\kappa}}\right) \| \tilde{x}^k - \tilde{y}^k \|^2 \leq 0$$

So $V_{k+1} \leq \rho^2 V_k$

i.e. $f(x^k) - f^* \leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)^k \underbrace{\left\{ f(x_0) - f^* + \frac{m}{2} \| x_0 - x^* \|^2 \right\}}_{= V_0}$

Convergence for Weakly Convex Functions

β_k is variable but $\alpha_k = \frac{1}{L}$

ρ_k

$$V_k = f(x^k) - f^* + \frac{L}{2} \| \tilde{x}^k - \rho_{k-1}^2 \tilde{x}^{k-1} \|^2$$

$m = 0$

$$V_{k+1} \leq \rho_k^2 V_k + R_k^{(weak)}$$

$$R_k^{(weak)} = \frac{L}{2} \| \tilde{y}^k - \rho_k^2 \tilde{x}^k \|^2 - \frac{\rho_k^2 L}{2} \| \tilde{x}^k - \rho_{k-1}^2 \tilde{x}^{k-1} \|^2$$

Want $R_k^{(weak)} = 0$ for $k \geq 1$

So $\tilde{y}^k - \rho_k^2 \tilde{x}^k = \rho_k \tilde{x}^k - \rho_k \rho_{k-1}^2 \tilde{x}^{k-1}$

Since $\tilde{y}^k = (1+\beta_k)\tilde{x}^k - \beta_k \tilde{x}^{k-1}$

we get $$\left.\begin{array}{l} 1+\beta_k - \rho_k^2 = \rho_k \\ \beta_k = \rho_k \rho_{k-1}^2 \end{array}\right\} \Rightarrow \rho_k^2 = \frac{(1-\rho_k^2)^2}{(1-\rho_{k-1}^2)^2}$$

$k = 1, 2, \ldots$

So $V_k \leq \rho_{k-1}^2 V_{k-1}$

$$\Rightarrow V_k \leq \rho_{k-1}^2 \rho_{k-2}^2 \cdots \rho_1^2 V_1 = \frac{(1-\rho_{k-1}^2)^2}{(1-\rho_0^2)^2} V_1$$

One can bound $V_1 \leq \frac{L}{2}\|x^0 - x^*\|^2$

So $V_k \leq (1-\rho_{k-1}^2)^2 \frac{L}{2}\|x^0 - x^*\|^2$

Inductive shows $1-\rho_k^2 \leq \frac{2}{k+2}$ (see text)

$$\boxed{f(x^k) - f^* \leq V_k \leq \frac{2L}{(k+1)^2}\|x^0 - x^*\|^2}$$ (Convergence rate) $\sim \frac{1}{t^2}$

Conjugate Gradient Method: Solve $Qx = p$

$$x^{k+1} = x^k - \alpha_k y^k$$

$$y^k = -\nabla f(x^k) + \beta_{k-1} y^{k-1}$$

Without L, m how to choose α_k and β_k?

$$\alpha_k = \operatorname{argmin}_{\alpha > 0} f(x^k + \alpha y^k)$$

when $f(x) = x^T Q x - p^T x + c$ we get

$$\alpha_k = \frac{(y^k)^T r^k}{(y^k)^T Q y^k}, \quad r^k = Qx^k - p$$

$$\langle y^k, Qy^{k-1} \rangle = \langle -r^{k-1} + \beta y^{k-1}, Qy^{k-1} \rangle$$

$$= \langle -r^{k-1}, Qy^{k-1} \rangle + \beta_k \langle y^{k-1}, Qy^{k-1} \rangle = 0$$

so $$\beta_k = \frac{\langle r^{k-1}, Qy^{k-1} \rangle}{\langle y^{k-1}, Qy^{k-1} \rangle}$$

Directions y^k are orthogonal to y^{k-1} w.r.t. $\langle u, v \rangle_Q = u^T Q v$

It does not have analysis if f is not quadratic.

Nesterov's method has optimal convergence rate achieved by the method among algorithms that make use of gradient information at the iterates x^k

One can design carefully an example for which NO method that makes use of all gradient observed up to and including iteration k (i.e. $\nabla f(x^i)$ $i=1,2,\ldots,k$) can produce a sequence $\{x^k\}$ better than Nesterov in its convergence rate

Appendix to Lecture 3

$$R_k := -\frac{m\rho^2}{2}\|\tilde{x}^k-\tilde{y}^k\|^2 - \frac{m(1-\rho^2)}{2}\|\tilde{y}^k\|^2$$

$$+ L(\tilde{u}^k)^T(\tilde{y}^k-\rho^2\tilde{x}^k) - \frac{L}{2}\|\tilde{u}^k\|^2$$

$$+ \frac{L}{2}\|\tilde{x}^{k+1}-\rho^2\tilde{x}^k\|^2 - \frac{\rho^2 L}{2}\|\tilde{x}^k-\rho^2\tilde{x}^{k-1}\|^2.$$

因为 $L(\tilde{u}^k)^T(\tilde{y}^k-\rho^2\tilde{x}^k) - \frac{L}{2}\|\tilde{u}^k\|^2$

$$= -\frac{L}{2}\left[-2(\tilde{u}^k)^T(\tilde{y}^k-\rho^2\tilde{x}^k) + \|\tilde{u}^k\|^2\right]$$

$$= -\frac{L}{2}\left[\|\tilde{y}^k-\rho^2\tilde{x}^k-\tilde{u}^k\|^2 - \|\tilde{y}^k-\rho^2\tilde{x}^k\|^2\right]$$

注意 $\tilde{y}^k-\tilde{u}^k = y^k-\tilde{u}^k-y^* = x^{k+1}-y^* = \tilde{x}^{k+1}$

所以

$$R_k = -\frac{m\rho^2}{2}\|\tilde{x}^k-\tilde{y}^k\|^2 - \frac{m(1-\rho^2)}{2}\|\tilde{y}^k\|^2$$

$$-\frac{L}{2}\left[\|\tilde{x}^{k+1}-\rho^2\tilde{x}^k\|^2 - \|\tilde{y}^k-\rho^2\tilde{x}^k\|^2\right]$$

$$+\frac{L}{2}\left[\|\tilde{x}^{k+1}-\rho^2\tilde{x}^k\|^2 - \rho^2\|\tilde{x}^k-\rho^2\tilde{x}^{k-1}\|^2\right]$$

$$= -\frac{m\rho^2}{2}\|\tilde{x}^k-\tilde{y}^k\|^2 - \frac{m(1-\rho^2)}{2}\|\tilde{y}^k\|^2$$

$$+\frac{L}{2}\left[\|\tilde{y}^k-\rho^2\tilde{x}^k\|^2 - \rho^2\|\tilde{x}^k-\rho^2\tilde{x}^{k-1}\|^2\right]$$

$$\frac{R_k}{L} = -\frac{\rho^2}{2k}\|\tilde{x}^k - \tilde{y}^k\|^2 - \frac{1-\rho^2}{2k}\|\tilde{y}^k\|^2$$
$$+\frac{1}{2}\|\tilde{y}^k - \rho^2\tilde{x}^k\|^2 - \frac{\rho^2}{2}\|\tilde{x}^k - \rho^2\tilde{x}^{k-1}\|^2$$

$$\frac{1-\rho^2}{k} = \frac{1}{k^{\frac{3}{2}}}$$

$$\tilde{y}^k = y^k - y^* = x^k + \beta(x^k - x^{k-1}) - y^*$$
$$= \tilde{x}^k + \beta(\tilde{x}^k - \tilde{x}^{k-1})$$

$$\tilde{x}^k - \tilde{y}^k = -\beta(\tilde{x}^k - \tilde{x}^{k-1})$$

$$\frac{R_k}{L} = -\frac{\rho^2}{2k}\|\tilde{x}^k - \tilde{y}^k\|^2 - \frac{1}{2k^{\frac{3}{2}}}\|\tilde{y}^k\|^2$$
$$+\frac{1}{2}\left\|(1-\rho^2)\tilde{x}^k + \beta(\tilde{x}^k - \tilde{x}^{k-1})\right\|^2 - \frac{\rho^2}{2}\|\tilde{x}^k - \rho^2\tilde{x}^{k-1}\|^2$$

$$\tilde{x}^{k-1} = \tilde{x}^k - \frac{1}{\beta}(\tilde{y}^k - \tilde{x}^k)$$

$$\tilde{x}^k - \rho^2\tilde{x}^{k-1} = \tilde{x}^k - \rho^2\tilde{x}^k + \frac{\rho^2}{\beta}(\tilde{y}^k - \tilde{x}^k)$$

$$\cancel{\rho^2 = 1 - \frac{1}{\sqrt{k}}\ \beta = \frac{1}{\sqrt{k}}} \qquad \frac{\rho^2}{\beta} = \frac{1-\frac{1}{\sqrt{k}}}{\frac{\sqrt{k}-1}{\sqrt{k}+1}} = \cancel{\frac{1}{\sqrt{k}(\sqrt{k}+1)}} = \frac{\frac{\sqrt{k}-1}{\sqrt{k}}}{\frac{\sqrt{k}-1}{\sqrt{k}+1}} = \frac{\sqrt{k}+1}{\sqrt{k}} = 1 + \frac{1}{\sqrt{k}}$$

$$\frac{R_k}{L} = -\frac{\rho^2}{2k}\|\tilde{x}^k - \tilde{y}^k\|^2 - \frac{1}{2k\sqrt{k}}\|\tilde{y}^k\|^2$$

$$+\frac{1}{2}\|\tilde{y}^k - \rho^2\tilde{x}^k\|^2 - \frac{\rho^2}{2}\|(1-\rho^2)\tilde{x}^k + (1+\frac{1}{\sqrt{k}})(\tilde{y}^k - \tilde{x}^k)\|^2$$

$$= -\frac{\rho^2}{2k}\|\tilde{x}^k - \tilde{y}^k\|^2 - \frac{1}{2k\sqrt{k}}\|\tilde{y}^k\|^2$$

$$+\frac{1}{2}\|(1-\rho^2)\tilde{x}^k + (\tilde{y}^k - \tilde{x}^k)\|^2$$

$$-\frac{\rho^2}{2}\|(1-\rho^2)\tilde{x}^k + (1+\frac{1}{\sqrt{k}})(\tilde{y}^k - \tilde{x}^k)\|^2$$

$$= -\frac{\rho^2}{2k}\|\tilde{x}^k - \tilde{y}^k\|^2 - \frac{1}{2k\sqrt{k}}\|\tilde{x}^k + (\tilde{y}^k - \tilde{x}^k)\|^2$$

$$+\frac{1}{2}\|(1-\rho^2)\tilde{x}^k + (\tilde{y}^k - \tilde{x}^k)\|^2$$

$$-\frac{\rho^2}{2}\|(1-\rho^2)\tilde{x}^k + (1+\frac{1}{\sqrt{k}})(\tilde{y}^k - \tilde{x}^k)\|^2$$

Coeffocient of $\|\tilde{x}^k\|^2$ is $-\frac{1}{2k\sqrt{k}} + \frac{(1-\rho^2)^2}{2} - \frac{(1-\rho^2)^2\rho^2}{2}$

$$= -\frac{1}{2k\sqrt{k}} + \frac{(1-\rho^2)^3}{2} = 0$$

Coeffocient of $(\tilde{x}^k)^T(\tilde{y}^k - \tilde{x}^k)$ is

$$-\frac{1}{k\sqrt{k}} + (1-\rho^2) - \frac{\rho^2}{2}\cdot 2(1+\frac{1}{\sqrt{k}})^{(1-\rho^2)} = -\frac{1}{k\sqrt{k}} + 1 - \rho^2 - \rho^2(1+\frac{1}{\sqrt{k}}) \times (1-\rho^2)$$

$1-\rho^2 = \frac{1}{\sqrt{k}}$ so

$$-\frac{1}{k\sqrt{k}}+\frac{1}{\sqrt{k}}-\left(1-\frac{1}{\sqrt{k}}\right)\left(1+\frac{1}{\sqrt{k}}\right)\frac{1}{\sqrt{k}}$$

$$=\frac{1}{\sqrt{k}}\left(1-\frac{1}{k}\right)-\left(1-\frac{1}{k}\right)\frac{1}{\sqrt{k}}=0$$

Coefficient of $\|\tilde{x}^k-\tilde{y}^k\|^2$ is

$$-\frac{\rho^2}{2k}-\frac{1}{2k\sqrt{k}}+\frac{1}{2}-\frac{\rho^2}{2}\left(1+\frac{1}{\sqrt{k}}\right)^2$$

$$=-\frac{\rho^2}{2k}-\frac{1}{2k\sqrt{k}}+\frac{1}{2}-\frac{\rho^2}{2}\left(1+\frac{2}{\sqrt{k}}+\frac{1}{k}\right)$$

$$=\frac{1}{2}-\frac{1}{2k\sqrt{k}}-\frac{\rho^2}{2}-\frac{\rho^2}{\sqrt{k}}-\frac{\rho^2}{k}$$

$$\|$$

$$\left(\rho^2=1-\frac{1}{\sqrt{k}}\right)\qquad \frac{1}{2}-\frac{1}{2k\sqrt{k}}-\frac{1-\frac{1}{\sqrt{k}}}{2}-\frac{1-\frac{1}{\sqrt{k}}}{\sqrt{k}}-\frac{1-\frac{1}{\sqrt{k}}}{k}$$

$$=\cancel{\frac{1}{2}}-\frac{1}{2k\sqrt{k}}\cancel{-\frac{1}{2}}+\frac{1}{2\sqrt{k}}-\frac{1}{\sqrt{k}}\cancel{+\frac{1}{k}}\cancel{-\frac{1}{k}}+\frac{1}{k\sqrt{k}}$$

$$=\frac{1}{2}\left(\frac{1}{k\sqrt{k}}-\frac{1}{\sqrt{k}}\right)$$

$$=-\frac{1}{2}\frac{1}{\sqrt{k}}\left(1-\frac{1}{k}\right)$$

$$=-\frac{1}{2}\frac{1}{\sqrt{k}}\left(1+\frac{1}{\sqrt{k}}\right)\rho^2 \qquad k=\frac{L}{m}$$

So $\dfrac{R_k}{L}=-\dfrac{1}{2}\rho^2\dfrac{1}{\sqrt{k}}\left(1+\dfrac{1}{\sqrt{k}}\right)\|\tilde{x}^k-\tilde{y}^k\|^2 \Rightarrow R_k=-\dfrac{1}{2}L\rho^2\dfrac{1}{\sqrt{k}}\left(1+\dfrac{1}{\sqrt{k}}\right)\times\|\tilde{x}^k-\tilde{y}^k\|^2$