20 elliptic functionals; iterative methods

Tuesday, November 17, 2020 3:49 AM

We are going to continue generalizing results to Hilbert spaces. Today, we will start with elliptic functionals, for which we can prove nice iterative convergence results

Def 13.7/6 Given a Hilbert space V, a functional $J: V \to \mathbb{R}$ is *elliptic* if it is continuously differentiable on V, and if there is some constant $\alpha > 0$ s.t.

$$\langle \nabla J_v - \nabla J_u, v-u \rangle \geq \alpha \|v-u\|^2 \quad \forall u, v \in V$$

Thm 13.6/13.8 Let V be a Hilbert space

(1) An elliptic functional $J: V \to \mathbb{R}$ is strictly convex and coercive. Furthermore, $J(v) - J(u) \geq \langle \nabla J_u, v-u \rangle + \frac{\alpha}{2}\|v-u\|^2 \quad \forall u, v \in V$.

(2) If $U \subseteq V$ nonempty, convex, and closed, and if J is an elliptic functional, then $u = \arg\min_{v \in U} J(v)$ is unique.

(3) Suppose $U \subseteq V$ is convex and J is elliptic. Then $u \in U$ is a minimum w.r.t. U iff

$$\langle \nabla J_u, v-u \rangle \geq 0 \quad \forall v \in U$$

in the general case, or

$$\nabla J_u = 0 \quad \text{if } U = V.$$

(4) A functional J which is twice differentiable in V is elliptic iff

$$\langle \nabla^2 J_u(w), w \rangle \geq \alpha \|w\|^2 \quad \forall u, w \in V.$$

proof. (1) We will use Taylor's formula with integral remainder:

$$\begin{aligned}
J(v) - J(u) &= \int_0^1 dJ_{u+t(v-u)}(v-u)\,dt \\
&= \int_0^1 \langle \nabla J_{u+t(v-u)}, v-u \rangle\,dt \\
&= \langle \nabla J_u, v-u \rangle + \int_0^1 \langle \nabla J_{u+t(v-u)} - \nabla J_u, v-u \rangle\,dt \\
&= \langle \nabla J_u, v-u \rangle + \int_0^1 \frac{1}{t} \langle \nabla J_{u+t(v-u)} - \nabla J_u, t(v-u) \rangle\,dt \\
&\geq \langle \nabla J_u, v-u \rangle + \int_0^1 \alpha t \|v-u\|^2\,dt \qquad (J \text{ is elliptic}) \\
&= \langle \nabla J_u, v-u \rangle + \frac{\alpha}{2}\|v-u\|^2.
\end{aligned}$$

Then $J(v) > J(u) + \langle \nabla J_u, v-u \rangle$, $\forall v \neq u$, so J is strictly convex.

Also, by Cauchy-Schwarz,

$$\begin{aligned}
J(v) &\geq J(0) + \langle \nabla J_0, v \rangle + \frac{\alpha}{2}\|v\|^2 \\
&\geq \underbrace{J(0) - \|\nabla J_0\|\,\|v\| + \frac{\alpha}{2}\|v\|^2}_{\to +\infty \text{ as } \|v\| \to +\infty.}
\end{aligned}$$

$\Rightarrow J$ is coercive. $\boxed{}$

(2) J is coercive and strictly convex, so it has a unique min. on a nonempty, convex, closed $U \subseteq V$.

(3) These are the conditions in Theorem 4.5 for part (2).

(4) Assume J is elliptic and twice differentiable.

$$D^2 J_u(w,w) = D_w\left[DJ_u(w)\right] = \lim_{\theta \to 0} \frac{DJ_{u+\theta w}(w) - DJ_u(w)}{\theta}$$

$$\Rightarrow \quad \langle \nabla^2 J_u(w), w \rangle = \lim_{\theta \to 0} \frac{\langle \nabla J_{u+\theta w} - \nabla J_u, w \rangle}{\theta}$$

$$= \lim_{\theta \to 0} \frac{1}{\theta^2} \langle \nabla J_{u+\theta w} - \nabla J_u, \theta w \rangle$$

$$\geq \alpha \|w\|^2.$$

Alternately, assume $\langle \nabla^2 J_u(w), w \rangle \geq \alpha \|w\|^2 \quad \forall u, w \in V.$

Define $g: V \to \mathbb{R}$ by $g(w) = \langle \nabla J_w, v-u \rangle = dJ_w(v-u) = D_{v-u} J(w)$. (fix $u, v \in V$)

Then $dg_{u+\theta(v-u)}(v-u) = D_{v-u}\, g(u+\theta(v-u)) = D_{v-u} D_{v-u} J(u+\theta(v-u))$

$$= D^2 J_{u+\theta(v-u)}(v-u, v-u).$$

Applying the Taylor-Maclaurin formula,

$$\langle \nabla J_v - \nabla J_u, v-u \rangle = g(v) - g(u)$$

$$= dg_{u+\theta(v-u)}(v-u) \qquad \text{(for some } 0<\theta<1\text{)}$$

$$= D^2 J_{u+\theta(v-u)}(v-u, v-u)$$

$$= \langle \nabla^2 J_{u+\theta(v-u)}(v-u), v-u \rangle$$

$$\geq \alpha \|v-u\|^2$$

$\Rightarrow J$ is elliptic. $\boxed{}$

$\boxed{}$

Corollary 13.1/13.9 If $J: \mathbb{R}^n \to \mathbb{R}$ is a quadratic function given by $J(v) = \frac{1}{2}\langle Av, v \rangle - \langle b, v \rangle$, where A is symmetric, then J is elliptic iff A is positive definite.

proof. $\langle \nabla^2 J_u(w), w \rangle = \langle Aw, w \rangle \geq \lambda_1 \|w\|^2$, where λ_1 is the smallest eigenval of A.

Similarly, $\langle \nabla^2 J_u(w), w \rangle \leq \lambda_n \|w\|^2$; λ_n the largest eigenval of A.

This leads us now to iterative methods on unconstrained problems

Similarly, $\langle J''_u(w), w \rangle \leq \lambda_n \|w\|^2$; λ_n the largest eigenval of A.

This leads us now to iterative methods on unconstrained problems, such as Newton's method, or things like Gauss–Seidel for solving $Ax=b$.

Sec. 13.5 Iterative methods for unconstrained problems

Recall: Given a desired solution u (e.g. a minimum of J, or u s.t. $Au=b$), an iterative method takes an initial vector u_0 and provides a rule for constructing a sequence $(u_k)_{k\geq 0}$ that converges to u.

Vol 1, Ch. 9: Wanted to solve $Au=b$, A invertible, $A=D-E-F$, where D diagonal, E below diagonal, F above diagonal.

Let $B = M^{-1}N$

Jacobi: Let $M = D \qquad N = E+F$

Gauss-Seidel: $M = D-E \qquad N = F$

Method of Relaxation: $M = \left(\frac{D}{\omega} - E\right) \qquad N = \left(\frac{1-\omega}{\omega} D + F\right)$, $0<\omega<2$

Then $u = Bu + M^{-1}b$, and if $\rho(B)<1$, then $\forall u_0$ $(u_k)_{k\geq 0}$, where $u_{k+1} = Bu_k + M^{-1}b$, converges to u.

Let's review Gauss Seidel in more detail:

$$u_{k+1} = (D-E)^{-1}Fu_k + (D-E)^{-1}b$$

$$\Rightarrow \quad Du_{k+1} = Eu_{k+1} + Fu_k + b$$

$$\begin{array}{lcccccccc}
a_{11}u_1^{k+1} & = & 0 & -a_{12}u_2^k & -a_{13}u_3^k & \cdots & -a_{1n}u_n^k & + b_1 \\
a_{22}u_2^{k+1} & = & -a_{21}u_1^{k+1} & 0 & -a_{23}u_3^k & \cdots & -a_{2n}u_n^k & + b_2 \\
\vdots & & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\
a_{nn}u_n^{k+1} & = & -a_{n1}u_1^{k+1} & -a_{n2}u_2^{k+1} & \cdots & -a_{n,n-1}u_{n-1}^{k+1} & 0 & + b_n
\end{array}$$

Back to optimization

Iterated line search

Consider the problem of finding $u = \underset{v\in V}{\arg\min}\, J(v)$, V a Hilbert space.

Optimizing over a Hilbert space is hard. But optimizing over a single variable is easier, so let's try that iteratively.

(1) At the current u_k, find a descent direction d_k, usually determined from the gradient of J at various pts. We need $\langle \nabla J_{u_k}, d_k \rangle < 0$.

(2) Exact line search: Consider the line $\{u_k + \mathbb{R}^+ d_k\}$

determined from the gradient of J at various pts. We need $\langle \nabla J_{u_k}, d_k \rangle < 0$.

(2) **Exact line search**: Consider the line $\{u_k + \mathbb{R}^+ d_k\}$.

Find $\rho_k \in \mathbb{R}^+$ s.t. $J(u_k + \rho_k d_k) = \inf_{\rho \in \mathbb{R}^+} J(u_k + \rho d_k)$

If ρ_k is unique, set $u_{k+1} = u_k + \rho_k d_k$. sometimes called "step-size"

(Also called line search or line minimization)

Prop 13.3/13.11 If J is a quadratic elliptic functional of the form
$$J(v) = \frac{1}{2} a(v,v) - h(v),$$
then given d_k, $\exists$ a unique ρ_k solving the line search.

proof. Lemma (Prop. 13.2/13.3): $J(u+\rho v) = \frac{\rho^2}{2} a(v,v) + \rho \langle \nabla J_u, v \rangle + J(u)$.

proof sketch: direct computation, and using $\nabla J_u = Au - b$. $\boxtimes$

Applying the Lemma, $J(u_k + \rho d_k) = \frac{\rho^2}{2} \underbrace{a(d_k, d_k)}_{\substack{>0 \text{ since} \\ J \text{ elliptic}}} + \rho \langle \nabla J_{u_k}, d_k \rangle + J(u_k)$.

Then $\frac{d}{d\rho}(u_k + \rho d_k) = \rho\, a(d_k, d_k) + \langle \nabla J_{u_k}, d_k \rangle = 0$ at the minimum.

$$\Rightarrow \quad \rho = \frac{-\langle \nabla J_{u_k}, d_k \rangle}{a(d_k, d_k)} > 0 \text{ is the minimum.} \quad \boxtimes$$

Unfortunately, step 2 (line search) is often too slow, so people have come up with alternatives, such as:

(3) **Backtracking line search**: Pick $\alpha, \beta \in \mathbb{R}$ s.t. $0 < \alpha < \frac{1}{2}$ and $0 < \beta < 1$, and set $t = 1$. Given a descent direction d_k at $u_k \in \mathrm{dom}(J)$,

while $J(u_k + t d_k) > J(u_k) + \alpha t \langle \nabla J_{u_k}, d_k \rangle$, **do** $t := \beta t$

Then set $\rho_k = t$; $u_{k+1} = u_k + \rho_k d_k$.

Instead of finding the actual minimum, we just try to decrease the value by "sufficiently much", and then pick a new direction. We start with a big jump, and then backtrack geometrically.

Cyclic coordinate descent (Method of relaxation)

Simplest method to pick directions in $\mathbb{R}^n$ in a cyclic fashion

Let $u_k = (u_1^k, u_2^k, \ldots, u_n^k)$

Then $J(u_1^{k+1}, u_2^k, \ldots, u_n^k) = \inf_{\lambda \in \mathbb{R}} J(\lambda, u_2^k, \ldots, u_n^k)$

$J(u_1^{k+1}, u_2^{k+1}, u_3^k, \ldots, u_n^k) = \inf_{\lambda \in \mathbb{R}} J(u_1^{k+1}, \lambda, u_3^k, \ldots, u_n^k)$

$$J(u_1^{k+1}, u_2^{k+1}, u_3^k, \dots, u_n^k) = \inf_{\lambda \in \mathbb{R}} J(u_1^{k+1}, \lambda, u_3^k, \dots, u_n^k)$$

$$\vdots$$

$$J(u_1^{k+1}, u_2^{k+1}, \dots, u_{n-1}^{k+1}, u_n^{k+1}) = \inf_{\lambda \in \mathbb{R}} J(u_1^{k+1}, \dots, u_{n-1}^{k+1}, \lambda)$$

If J is differentiable and convex, then necessary and sufficient conditions are that $dJ_v(e_i) = 0$ for $i = 1, \dots, n$

$$\iff \langle \nabla J_v, e_i \rangle = 0$$

Prop. 13.4 If the functional $J : \mathbb{R}^n \to \mathbb{R}$ is elliptic, then the relaxation method converges.

Aside: If $J(v) = \frac{1}{2} v^T A v - b^T v$, where A is symm. pos. def.,

then $\nabla J_v = Av - b = 0$ for v that minimize J.

Turns out that using the "method of relaxation" on minimizing J is equivalent to Gauss-Seidel on the system $Av = b$.

13.6 Gradient descent

A moment ago we chose to cycle through each coordinate to determine which direction to go. What are other obvious choices?

Recall: Needed $\langle \nabla J_{u_k}, d_k \rangle < 0$. What if we use $d_k = -\nabla J_{u_k}$ instead?

Note $J(u_k + w) = J(u_k) + \langle \nabla J_{u_k}, w \rangle + \varepsilon(w)\|w\|$, with $\lim_{w \to 0} \varepsilon(0) = 0$. (Taylor's Thm)

So if $\nabla J_{u_k} \neq 0$, the 1st order part of the variation in J is bounded by $\|\nabla J_{u_k}, w\|\|w\|$ by Cauchy-Schwarz,

and equality achieved if ∇J_{u_k} and w are colinear.

Gradient descent iteration: $u_{k+1} = u_k - \rho_k \nabla J_{u_k}$, where $\rho_k > 0$.

Choosing ρ_k:

(1) Gradient method with fixed stepsize parameter. $\rho_k = \rho$, a constant for all k.
(2) Gradient method with variable stepsize parameter. ρ_k is variable.
(3) Gradient method with optimal stepsize. Also called steepest descent for the Euclidean norm. Choose ρ_k by the line search

$$J(u_k - \rho_k \nabla J_{u_k}) = \inf_{\rho \in \mathbb{R}} J(u_k - \rho \nabla J_{u_k})$$

$$J(u_k - \rho_k \nabla J_{u_k}) = \inf_{\rho \in \mathbb{R}} J(u_k - \rho \nabla J_{u_k})$$

As in other optimal line searches only works if always a unique min.

(4) Gradient descent with backtracking line search.

Prop 13.5/13.3 Let $J : \mathbb{R}^n \to \mathbb{R}$ be an elliptic functional. Then the gradient method with optimal stepsize parameter converges.

proof. Since J is elliptic, J has a unique min. at $\nabla J_u = 0$.

NTS that the sequence $(u_k)_{k \geq 0}$ converges to u, from any start u_0.
WLOG, $u_{k+1} \neq u_k$ and $\nabla J_{u_k} \neq 0$, or else we're done after finite steps.

Lemma 1: Any two consecutive descent directions are orthogonal and
$$J(u_k) - J(u_{k+1}) \geq \frac{\alpha}{2} \|u_k - u_{k+1}\|^2.$$

proof. Let $\varphi_k : \mathbb{R} \to \mathbb{R}$ be defined by $\varphi_k(\rho) = J(u_k - \rho \nabla J_{u_k})$, which is strictly convex and coercive.

So φ_k has a unique min ρ_k at $\varphi_k'(\rho) = 0$.

By the chain rule,
$$\varphi_k'(\rho) = dJ_{u_k - \rho \nabla J_{u_k}}(-\nabla J_{u_k})$$
$$= -\langle \nabla J_{u_k - \rho \nabla J_{u_k}}, \nabla J_{u_k} \rangle$$
$(u_{k+1} = u_k - \rho \nabla J_{u_k})$
$$= -\langle \nabla J_{u_{k+1}}, \nabla J_{u_k} \rangle = 0.$$
(proved orthogonality)

Furthermore, $\langle \nabla J_{u_{k+1}}, u_{k+1} - u_k \rangle$
$$= \langle \nabla J_{u_{k+1}}, -\rho \nabla J_{u_k} \rangle = 0$$

Then by Thm 13.6/13.8,
$$J(u_k) - J(u_{k+1}) \geq \langle \nabla J_{u_{k+1}}, u_{k+1} - u_k \rangle + \frac{\alpha}{2} \|u_k - u_{k+1}\|^2 = \frac{\alpha}{2} \|u_k - u_{k+1}\|^2. \quad \boxtimes$$

Lemma 2: $\lim_{k \to \infty} \|u_k - u_{k+1}\| = 0.$

proof. Clearly, by Lemma 1, $(J(u_k))_{k \geq 0}$ is decreasing and bounded below by $J(u)$.

Therefore, the sequence must converge to something
$$\Rightarrow \lim_{k \to \infty} (J(u_k) - J(u_{k+1})) = 0$$
$$\Rightarrow \lim_{k \to \infty} \|u_k - u_{k+1}\| = 0. \quad \boxtimes$$

Lemma 3: $\|\nabla J_{u_k}\| \leq \|\nabla J_{u_k} - \nabla J_{u_{k+1}}\|.$

Lemma 3: $\|\nabla J_{u_k}\| \le \|\nabla J_{u_k} - \nabla J_{u_{k+1}}\|$.

proof. Consecutive descent directions are orthogonal, so by Cauchy Schwarz,

$$\|\nabla J_{u_k}\|^2 = \langle \nabla J_{u_k}, \nabla J_{u_k} - \nabla J_{u_{k+1}} \rangle$$
$$\le \|\nabla J_{u_k}\| \, \|\nabla J_{u_k} - \nabla J_{u_{k+1}}\|$$
$$\Rightarrow \|\nabla J_{u_k}\| \le \|\nabla J_{u_k} - \nabla J_{u_{k+1}}\| \qquad \boxed{\diagup}$$

Lemma 4: $\lim_{k\to\infty} \|\nabla J_{u_k}\| = 0$.

proof. Because $(J(u_k))_{k\ge 0}$ is decreasing, and J is coercive, $(u_k)_{k\ge 0}$ must be bounded.

By hypothesis, dJ is continuous so it is uniformly continuous over compact subsets of $\mathbb{R}^n$ (because $\mathbb{R}^n$ is finite dim)

Thus, $\forall \varepsilon > 0$, if $\|u_k - u_{k+1}\| < \varepsilon$, then

$$\|dJ_{u_k} - dJ_{u_{k+1}}\|_2 < \varepsilon$$ induced operator norm.

But $$\|dJ_{u_k} - dJ_{u_{k+1}}\|_2 = \sup_{\|w\|\le 1} |dJ_{u_k}(w) - dJ_{u_{k+1}}(w)|$$ (operator norm def)
$$= \sup_{\|w\|\le 1} |\langle \nabla J_{u_k} - \nabla J_{u_{k+1}}, w\rangle|$$
$$\le \|\nabla J_{u_k} - \nabla J_{u_{k+1}}\|.$$ (Cauchy-Schwarz)

And $$\|\nabla J_{u_k} - \nabla J_{u_{k+1}}\|^2 = \langle \nabla J_{u_k} - \nabla J_{u_{k+1}}, \nabla J_{u_k} - \nabla J_{u_{k+1}} \rangle$$
$$= dJ_{u_k}(\nabla J_{u_k} - \nabla J_{u_{k+1}}) - dJ_{u_{k+1}}(\nabla J_{u_k} - \nabla J_{u_{k+1}})$$
$$\le \|dJ_{u_k} - dJ_{u_{k+1}}\|_2 \, \|\nabla J_{u_k} - \nabla J_{u_{k+1}}\|$$

$$\Rightarrow \|dJ_{u_k} - dJ_{u_{k+1}}\|_2 = \|\nabla J_{u_k} - \nabla J_{u_{k+1}}\|.$$

Thus, $$\lim_{k\to\infty} \|\nabla J_{u_k} - \nabla J_{u_{k+1}}\| = \lim_{k\to\infty} \|dJ_{u_k} - dJ_{u_{k+1}}\|_2 = 0$$

since $\lim_{k\to\infty} \|u_k - u_{k+1}\| = 0$ (Lem 2) and dJ is continuous

And by Lem 3, $\|\nabla J_{u_k}\| \le \|\nabla J_{u_k} - \nabla J_{u_{k+1}}\|$

$$\Rightarrow \lim_{k\to\infty} \|\nabla J_{u_k}\| = 0. \qquad \boxed{\diagup}$$

Lemma 5: The sequence $(u_k)_{k\ge 0}$ converges to the minimum u.

proof. J is elliptic and $\nabla J_u = 0$, so

Lemma 5: The sequence $(u_k)_{k\ge 0}$ converges to the minimum u.

proof. J is elliptic and $\nabla J_u = 0$, so

$$\alpha \|u_k - u\|^2 \le \langle \nabla J_{u_k} - \nabla J_u, u_k - u\rangle$$
$$= \langle \nabla J_{u_k}, u_k - u\rangle$$
$$\le \|\nabla J_{u_k}\| \, \|u_k - u\|$$

$$\Rightarrow \|u_k - u\| \le \frac{1}{\alpha}\|\nabla J_{u_k}\|$$

$$\Rightarrow \lim_{k\to\infty} \|u_k - u\| \le \frac{1}{\alpha}\lim_{k\to\infty}\|\nabla J_{u_k}\| = 0 \qquad \blacksquare$$

For an elliptic functional $J: \mathbb{R}^n \to \mathbb{R}$, the optimal step-size method can be given as follows:

Note $J(v) = \frac{1}{2}\langle Av, v\rangle - \langle b, v\rangle$ and $\nabla J_v = Av - b$, so

$$0 = \langle \nabla J_{u_{k+1}}, \nabla J_{u_k}\rangle = \langle A(u_k - \rho_k(Au_k - b)) - b, Au_k - b\rangle$$

$$\Rightarrow \rho_k = \frac{\|w_k\|^2}{\langle Aw_k, w_k\rangle}, \qquad w_k = Au_k - b = \nabla J_{u_k}.$$

So we get steps:

(1) $w_k = Au_k - b$

(2) $\rho_k = \dfrac{\|w_k\|^2}{\langle Aw_k, w_k\rangle}$

(3) $u_{k+1} = u_k - \rho_k w_k$.

fast especially if Aw is cheap to compute e.g. if A is sparse.

Convergence proofs for other variants of gradient descent are available, though they may require more or fewer conditions. e.g. see 13.7 for convergence of gradient descent with variable step-sizes in (infinite-dim) Hilbert spaces where J is elliptic and the gradient of J is Lipschitz. (Prop. 13.6)

Alternately, can also do steepest descent w.r.t. another norm (instead of Euclidean 2-norm)

13.10 Conjugate gradients for unconstrained optimization

One more method for unconstrained optimizations.

Intuition is that setting descent $d_k = -\nabla J_{u_k}$ is not always optimal.

Consider ordinary gradient descent. We have $\langle \nabla J_{u_k}, \nabla J_{u_{k+1}}\rangle = 0$ for consecutive directions. But you might "repeat" directions during a long run. Can we ensure that

Consider ordinary gradient descent. We have $\langle \nabla J_{u_k}, \nabla J_{u_{k+1}} \rangle = 0$ for consecutive directions.
But you might "repeat" directions during a long run. Can we ensure that we never repeat directions?

Normally, given u_k, $u_k = u_k - \rho_k \nabla J_{u_k}$, where $\rho_k = \inf_{\rho \in \mathbb{R}} J(u_k - \rho \nabla J_{u_k})$

Instead, let's try minimizing J over $u_k + \mathcal{G}_k$, where $\mathcal{G}_k = \text{span}\{\nabla J_{u_0}, \dots, \nabla J_{u_k}\} \subseteq \mathbb{R}^n$.

i.e. Find $u_{k+1} \in u_k + \mathcal{G}_k$ and $J(u_{k+1}) = \inf_{v \in u_k + \mathcal{G}_k} J(v)$.

Has several nice properties, assuming J is elliptic.

(1) The gradients ∇J_{u_i} and ∇J_{u_j} are orthogonal $\forall i,j$ with $0 \le i < j \le k$.

Thus, if $\nabla J_{u_i} \neq 0$ for $i = 0, \dots, k$, then $\{\nabla J_{u_i}\}_{i=0,\dots,k}$ are lin. ind., so the method terminates in at most n steps.

(2) Let $\Delta_\ell = u_{\ell+1} - u_\ell = -\rho_\ell d_\ell$. Then $\langle A\Delta_\ell, \Delta_i \rangle = 0$ $\quad 0 \le i < \ell \le k$.

i.e. Δ_ℓ and Δ_i are "A-conjugate." Thus, if $\Delta_\ell \neq 0$ for $\ell = 0, \dots, k$, then Δ_ℓ are lin. ind.

(3) There is a simple formula to compute d_{k+1} from d_k and to compute ρ_k.

If $\nabla J_{u_i} \neq 0$ for $i = 0, \dots, k$, then we can write

$$d_\ell = \sum_{i=0}^{\ell-1} \lambda_i^\ell \nabla J_{u_i} + \nabla J_{u_\ell}, \quad 0 \le \ell \le k.$$

Then

$$\begin{cases} \lambda_i^k = \dfrac{\|\nabla J_{u_k}\|^2}{\|\nabla J_{u_i}\|^2}, & 0 \le i \le k-1 \\ d_0 = \nabla J_{u_0} & \\ d_\ell = \nabla J_{u_\ell} + \dfrac{\|\nabla J_{u_\ell}\|^2}{\|\nabla J_{u_{\ell-1}}\|^2} d_{\ell-1}, & 1 \le \ell \le k. \end{cases}$$

We will not prove these properties in class due to time constraints.