Solving linear systems: gradient descent

Perhaps the first physical application of solving a linear system of equations one comes across is in the study of resistive circuits.

Consider a circuit represented as a weighted undirected graph G, where the weights on the edges represent their conductances. If current i_a is being fed in externally at a node a, then Kirchoff's current law (charge balance) takes the form

$$\sum_{b \sim a} c_{ab} (V_a - V_b) = i_a$$

where $b \sim a$ denotes "b adjacent to a", and V_a is the voltage at node a.

c_{ab} is the conductance of the edge a-b.

This system can be written down as

$$LV = I$$

where I is the vector of external current inputs, v is the voltage vector, and L, the Laplacian is a matrix given as

$$\left.\begin{aligned} L_{aa} &= \sum_{b\sim a} c_{ab} \\ L_{ab} = L_{ba} &= -c_{ab}. \end{aligned}\right\} \text{L is symmetric}$$

Now we note that for any vector v,

$$v^T L v = \sum_{\substack{e\in E \\ e=(a,b)}} c_{ab}\,(v_a - v_b)^2 \geq 0$$

So, L is symmetric positive semi-definite (P.S.D.)

We are thus led to study linear systems of the form

$Ax = b$ where $A \succeq 0$ i.e. A is positive semi definite.

In fact if A is not <u>strictly</u> positive definite (i.e., if $\exists v$ s.t. $v^TAv=0$), then $Ax=b$ does not have a unique solution. So we will make the stronger assumption that

$$v^TAv > 0 \text{ for all } v$$

<u>Note</u>:- This assumption is <u>not</u> satisfied by the <u>Laplacian</u>: $L\vec{1}=0$.

However, when the graph is connected, the nullspace of L is spanned by $\vec{1}$, and one can convert it into a problem of the form $Ax=b$ with A PSD. This will be a homework problem.

As we saw briefly in lecture 1, gradient descent can be used to solve such "root finding" tasks. For real valued fns (such as $f(x) = x^2 - q$ in that lecture) this works by finding a fn. g such that

$$g'(x) = f(x),$$

and then using an iterative procedure of the form:

For $k = 0$ to t,

$$x_{k+1} = x_k - \varepsilon(x_k)\, g'(x_k)$$

for some appropriate step sizes

$$\varepsilon(x_k) \geq 0.$$

In this setting we want to solve

$$f(\vec{x}) = 0 \text{ where } f(\vec{x}) = Ax - b.$$

We consider

$$g(\vec{x}) = \frac{1}{2} x^T A x - b^T x$$

We verify that

$$\nabla g(\vec{x}) = \frac{1}{2}(A + A^T)\, x - b,$$

So that

$$\nabla g(\vec{x}) = Ax - b = f(x) \text{ when } A = A^T.$$

Note also that when A is positive definite

$\tilde{x} = A^{-1}b$ is the unique minimum of g. In fact, if we define the following inner product: $\langle u, v\rangle_A := u^T A v$ [Ex. verify that this is an inner product when A is symmetric positive definite], then

$$g(x) = \frac{1}{2}\|x - \tilde{x}\|_A^2 + g(\tilde{x}), \quad (*)$$

where

$$\|x - \tilde{x}\|_A^2 := (x-\tilde{x})^T A (x - \tilde{x}) > 0 \text{ when } x \neq \tilde{x}.$$

Proof:- $g(x) = \frac{1}{2}\|x\|_A^2 - x^T b$

$$= \frac{1}{2}\|x\|_A^2 - x^T A\tilde{x} \qquad (\because A\tilde{x} = b)$$

$$= \frac{1}{2}\|x\|_A^2 - \langle x, \tilde{x}\rangle_A$$

$$= \frac{1}{2}\|x - \tilde{x}\|_A^2 - \frac{1}{2}\|\tilde{x}\|_A^2$$

Thus $g(\tilde{x}) = -\frac{1}{2}\|\tilde{x}\|_A^2$ and $g(x) = \frac{1}{2}\|x-\tilde{x}\|_A^2 + g(\tilde{x})$. ∎

Note:- In the vicinity of its unique minimum, any smooth strictly convex function h has a similar form as $(*)$. More formally, in the one dimensional setting one has

$$h(x) = h(x_0) + h'(x_0)(x-x_0) + \frac{1}{2} h''(x_0)(x-x_0)^2 + O((x-x_0)^3),$$

by taylor expansion around the minimum x_0. $h'(x_0)=0$ (optimality condition) so we get

$$h(x) = h(x_0) + \frac{1}{2} h''(x_0)(x-x_0)^2 + O((x-x_0)^3),$$

where $h''(x_0) > 0$, by the assumption of strict convexity.

Similarly in higher dimensions:

$$h(\vec{x}) = h(\vec{x}_0) + \frac{1}{2} \|\vec{x} - \vec{x}\|_H^2 + O(\|x-x_0\|^3),$$

where H is the Hessian of h at $\vec{x}_0$.

Gradient descent for optimizing

$$g(x) = \frac{1}{2} x^T A x - b x.$$

Given current estimate x_k, we set

$$x_{k+1} = x_k - \varepsilon(x_k) \nabla g(x_k),$$

where $\varepsilon(x_k)$ is chosen as the minimum of $f(t) = g(x_k - t \nabla g(x_k))$.

Define $r_k := \nabla g(x_k) = A x_k - b$.

Then
$$f(t) = g(x_k - t r_k),$$
$$\begin{aligned}
f'(t) &= -(\nabla g(x_k - t r_k))^T r_k \\
&= -(A(x_k - t r_k) - b)^T r_k \\
&= (r_k - t A r_k)^T r_k \\
&= r_k^T r_k - t\, r_k^T A r_k \\
&= \|r_k\|^2 - t \|r_k\|_A^2
\end{aligned}$$

Thus $f'(t) = 0$ when $t = \dfrac{\|r_k\|^2}{\|r_k\|_A^2}$.

[Ex: this is also a minimum]. Thus the step is

$$x_{k+1} = x_k - \frac{\|r_k\|^2}{\|r_k\|_A^2} r_k \quad \text{where } r_k = A x_k - b$$

Note that each steps

① Does $O(n^2)$ arithmetic operations, only one of which is a divison (in the computation of $\frac{\|r_k\|^2}{\|r_k\|_A^2}$) while the rest are multiplication and addition.

② Only matrix-vector multiplications and inner products are performed. <u>No matrix inversion is performed</u>.