6.5.6 Conjugate Gradient Method (nonlinear)

- alternative to Newton's Method
- no 2nd derivatives (or approximations to them)
- steepest descent repeats search directions - inefficient
- CG ~~removes~~ modifies gradient to remove previous search directions

CG for unconstrained minimization

x_0 = initial guess

$g_0 = \nabla f(x_0)$

$s_0 = -g_0$

for $k = 0, 1, 2, \ldots$

choose α_k to minimize $f(x_k + \alpha_k s_k)$ [1D line search]

$x_{k+1} = x_k + \alpha_k s_k$ [update soln]

$g_{k+1} = \nabla f(x_{k+1})$ [compute new gradient]

(Fletcher & Reeves)

$\beta_{k+1} = g_{k+1}^T g_{k+1} / g_k^T g_k$

$s_{k+1} = -g_{k+1} + \beta_{k+1} s_k$ [modify gradient]

end.

alternative (Polak & Ribiere)

$$\beta_{k+1} = (g_{k+1} - g_k)^T g_{k+1} / g_k^T g_k$$

- F&R equivalent to P&R for quadratic function w/ exact line search
- P&R better for general non-linear functions or inexact line search.

- theoretically, ~~CG~~ CG exact after n-iterations for quad function in n-dim
- effective in general as well

- common to restart after n iterations
  - reinit to -gradient

Example 6.14 CG

$$f(x) = \frac{1}{2}x_1^2 + \frac{5}{2}x_2^2$$

$$\nabla f(x) = \begin{pmatrix} x_1 \\ 5x_2 \end{pmatrix}$$

$$x_0 = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$$

$$s_0 = -g_0 = -\nabla f(x_0) = -\begin{pmatrix} 5 \\ 5 \end{pmatrix}$$

Exact line search

$$0 = f'(x_0 + \alpha_0 s_0) = \nabla f(x_0 + \alpha_0 s_0)^T s_0 = *$$

$$x_0 + \alpha_0 s_0 = \begin{pmatrix} 5 \\ 1 \end{pmatrix} + \alpha_0 \begin{pmatrix} -5 \\ -5 \end{pmatrix} = \begin{pmatrix} 5 - 5\alpha_0 \\ 1 - 5\alpha_0 \end{pmatrix}$$

$$\nabla f(x_0 + \alpha_0 s_0) = \begin{pmatrix} 5 - 5\alpha_0 \\ 5(1 - 5\alpha_0) \end{pmatrix}$$

$$\nabla f(x_0 + \alpha_0 s_0)^T s_0 = \begin{pmatrix} 5 - 5\alpha_0 \\ 5(1 - 5\alpha_0) \end{pmatrix} \cdot \begin{pmatrix} +5 \\ +5 \end{pmatrix} = 0$$

$$\Rightarrow \quad 5 - 5\alpha_0 + 5 - 25\alpha_0 = 0$$

$$\Rightarrow \quad 10 = 30\alpha_0 \quad \Rightarrow \quad \boxed{\alpha_0 = 1/3}$$

$$x_1 = x_0 + \alpha_0 s_0 = \begin{pmatrix} 5 \\ 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} -5 \\ -5 \end{pmatrix} = \begin{pmatrix} +\frac{10}{3} \\ -\frac{2}{3} \end{pmatrix}$$

$$g_1 = \nabla f(x_1) = \begin{pmatrix} \frac{10}{3} \\ -\frac{10}{3} \end{pmatrix}$$ ← S.D. would now search along $-g_1$

$$\beta_1 = g_1^T g_1 / g_0^T g_0 = (100/9 + 100/9)/50 = \frac{200/50}{9} = \frac{4}{9}$$

$$s_1 = -g_1 + \beta_1 s_0 = \begin{pmatrix} -\frac{10}{3} \\ \frac{10}{3} \end{pmatrix} + \frac{4}{9}\begin{pmatrix} -5 \\ -5 \end{pmatrix} = \begin{pmatrix} -\frac{10}{3} - \frac{20}{9} \\ \frac{10}{3} - \frac{20}{9} \end{pmatrix} = \begin{pmatrix} -\frac{30}{9} - \frac{20}{9} \\ \frac{30}{9} - \frac{20}{9} \end{pmatrix} = \begin{pmatrix} -\frac{50}{9} \\ \frac{10}{9} \end{pmatrix} \rightarrow$$

min along that direction is @ $\alpha_1 = 0.6 = \frac{3}{5}$

$\leadsto$ gives

$$X_2 = X_1 + \alpha_1 S_1$$

$$= \begin{pmatrix} \frac{10}{3} \\ -\frac{2}{3} \end{pmatrix} + \frac{3}{5} \begin{pmatrix} -\frac{50}{9} \\ \frac{10}{9} \end{pmatrix} = \begin{pmatrix} \frac{10}{3} - \frac{\cancel{3}\cdot \overset{10}{\cancel{50}}}{\cancel{5}\cdot \cancel{9}3} \\ -\frac{2}{3} + \frac{\cancel{3}}{\cancel{5}} \cdot \frac{\cancel{10}2}{\cancel{9}3} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \checkmark$$

exact solution

---

converged in 2 steps.

6.2.3 Constrained Optimality Conditions

- for constrained optimization, solution often occurs on the boundary of the feasible set.

- principles the same: min at x^* when there is no downhill <u>feasible</u> direction

---

$\vec{s}$ is a <u>feasible direction</u> at $\vec{x}^* \in S$ (feasible set)

if $\exists\, r > 0$ s.t. $\vec{x}^* + \alpha \vec{s} \in S \;\; \forall\, \alpha \in [0, r]$

1st order necessary condition for optimality

For any feasible direction $\vec{s}$,

$$\boxed{\nabla f(x^*)^T \vec{s} \geq 0}$$

i.e. f doesn't decrease around x^* in any feasible direction

<u>Note</u>: at interior x^*, every s is feasible, so must have both $\nabla f(x^*)^T s \geq 0$, and

$\nabla f(x^*)^T(-s) \geq 0$

$\Rightarrow \boxed{\nabla f(x^*) = 0}$

1st order n.c. for interior (unconstrained) min.

2nd order necessary/sufficient condition for optimality

For any feasible direction $\vec{s}^*$,

$$\underbrace{s^T H_f(x^*) s \geq 0}_{\text{necessary}}, \quad \underbrace{s^T H_f(x^*) s > 0}_{\text{sufficient}}$$

Lagrange Multipliers

$$\min_x f(\vec{x})$$
$$\text{subj. to} \quad \vec{g}(\vec{x}) = \vec{0}$$

$$f: \mathbb{R}^n \to \mathbb{R}$$
$$g: \mathbb{R}^n \to \mathbb{R}^m, \qquad m \leq n$$

Necessary condition for feasible point x^* to be a solution is (for some $\lambda^* \in \mathbb{R}^m$, $J_g(x^*)$ Jacobian matrix of g)

$$-\nabla f(x^*) = J_g^T(x^*)\lambda^*$$

i.e., negative gradient of f lies in space spanned by <u>constraint normals</u>

λ^* are <u>Lagrange multipliers</u>.

i.e. cannot reduce f w/out violating constraints

Lagrangian Function $\quad \mathcal{L}: \mathbb{R}^{n+m} \to \mathbb{R}$

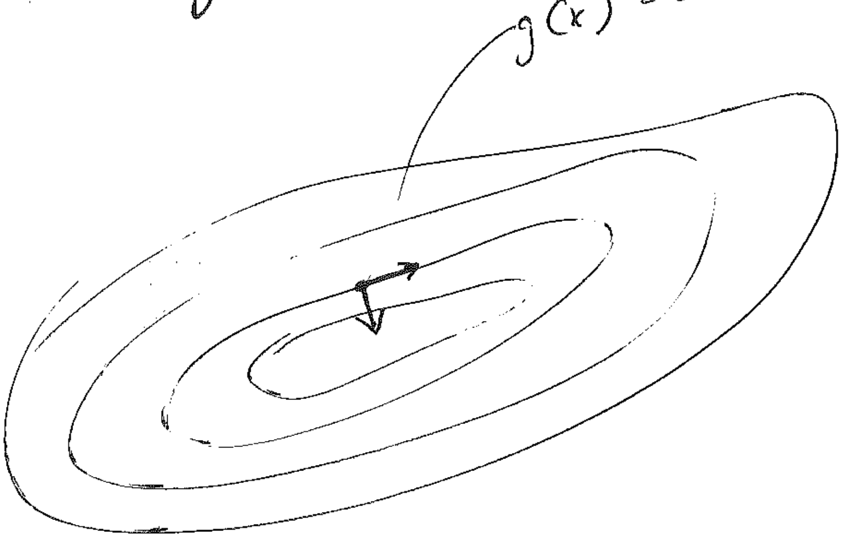
$$\mathcal{L}(x,\lambda) = f(x) + \lambda^T g(x)$$

$$\nabla \mathcal{L}(x,\lambda) = \begin{pmatrix} \nabla_x \mathcal{L}(x,\lambda) \\ \nabla_\lambda \mathcal{L}(x,\lambda) \end{pmatrix} = \begin{pmatrix} \nabla f(x) + J_g(x)^T \lambda \\ g(x) \end{pmatrix}$$

$$H_{\mathcal{L}}(x,\lambda) = \begin{pmatrix} B(x,\lambda) & J_g(x)^T \\ J_g(x) & 0 \end{pmatrix}$$

$$B(x,\lambda) = H_f(x) + \sum^m \lambda_i H_{g_i}(x)$$

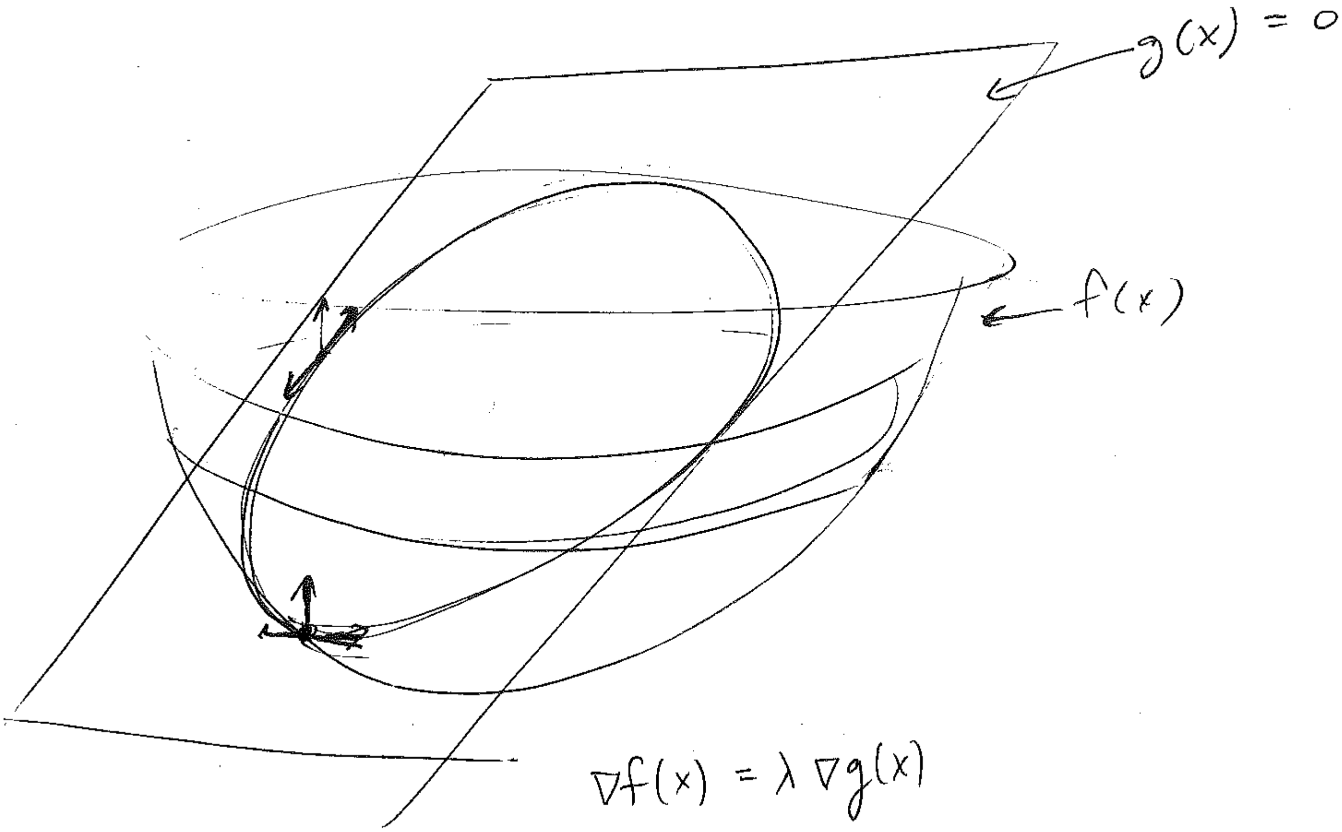
$g(x) = 0$



~~[illegible]~~ function g constant along $\vec{s}$

$\Rightarrow \nabla g(x)^T \vec{s} = 0$

i.e. gradient orthogonal to $\vec{s}$



$$\nabla f(x) = \lambda \nabla g(x)$$

Critical Point of Lagrangian

$m+n$ equations in $m+n$ unknowns

$$\nabla \mathcal{L}(x,\lambda) = \begin{pmatrix} \nabla f(x) + J_g^T(x)\lambda \\ g(x) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \quad -\nabla f(x^*) = J_g^T(x^*)\lambda, \quad \text{and} \quad \text{(necessary condition)}$$

$$g(x^*) = 0 \quad \text{(feasibility)}$$

critical pt of $\mathcal{L}$ is a saddle, not a minimum.

$H_{\mathcal{L}}(x,\lambda)$ is indefinite.

Solution x^* is a critical pt. of Lagrangian, but not a minimum.

Q. So how to check for optimality?

sufficient condition:

$B(x^*, \lambda^*)$ is pos. def. on tangent space to constraint surface, i.e. the null space of $J_g(x^*)$

Let Z be a matrix whose columns ~~span~~ form a basis for nullspace of $J_g(x^*)$. Check that

$$Z^T B Z$$ (projected, or reduced Hessian)

is positive definite

... factorization of $J_g(x^*)$ to obtain Z.

Example 6.6 (*) Equality Constrained Optimization

$$\mathcal{L}(x,\lambda) = f(x_1,x_2) + \lambda g(x_1,x_2)$$

$$f(x_1,x_2) = 2\pi x_1(x_1+x_2)$$
$$g(x_1,x_2) = \pi x_1^2 x_2 - V$$

$$\nabla f(x_1,x_2) = 2\pi\begin{pmatrix}2x_1+x_2\\ x_1\end{pmatrix}, \quad J_g(x) = \pi\begin{pmatrix}2x_1x_2 & x_1^2\end{pmatrix}$$

$$\Rightarrow \nabla\mathcal{L}(x,\lambda) = \begin{pmatrix}\nabla f(\vec{x}) + \lambda\nabla g(\vec{x})\\ g(x_1,x_2)\end{pmatrix} = \begin{pmatrix}2\cancel{\pi}(2x_1+x_2) + 2\lambda\cancel{\pi}x_1x_2\\ 2\cancel{\pi}x_1 + \lambda\cancel{\pi}x_1^2\\ \cancel{\pi}x_1^2x_2 - V/\pi\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

(*) $n=2$, $m=1$, x_1 = radius, x_2 = height
minimize surface area of a cylinder subject to constraint on volume:

$$\min_x f(x_1,x_2) = 2\pi x_1x_2 + 2\pi x_1^2 = 2\pi x_1(x_1+x_2)$$

$$\text{subj. to} \quad g(x_1,x_2) = \pi x_1^2x_2 - V = 0$$

Let $V = 1000\text{ cm}^3$, solve approximately $x_1 = 5.4\text{ cm}$, $x_2 = 10.8\text{ cm}$, $\lambda = -0.37$

Confirm optimality:

$$H_f(x) = 2\pi\begin{pmatrix}2 & 1\\ 1 & 0\end{pmatrix}, \text{ and } H_g(x) = 2\pi\begin{pmatrix}x_2 & x_1\\ x_1 & 0\end{pmatrix}$$

$$B(x^*,\lambda^*) = 2\pi\begin{pmatrix}2+\lambda x_2 & 1+\lambda x_1\\ 1+x_1\lambda & 0\end{pmatrix} = \begin{pmatrix}-12.6 & -6.3\\ -6.3 & 0\end{pmatrix}$$

$\lambda = -15.2,\ 2.6$ (not pos. def.)

$$J_g(x^*) = (369 \quad 92.3) \Rightarrow z = (-.243,\ 0.970)^T$$

$$z^TBz = 2.23 > 0 \quad \checkmark$$

$$f(x^*) = 554\text{ cm}^2 \quad \text{surface area.}$$