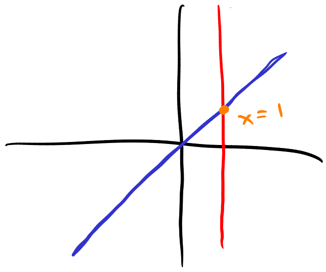
Augmented Lagrange

Simple example

$$\underset{x}{\text{minimize}} \quad ax$$
$$\text{s.t.} \quad x = 1$$



$\nearrow$ $h(x) = x - 1$

$$\nabla_x \mathcal{L}\Big|_{x^*,\lambda^*} = \nabla f(x^*) + \lambda^* \underbrace{\nabla h(x^*)} = 0$$

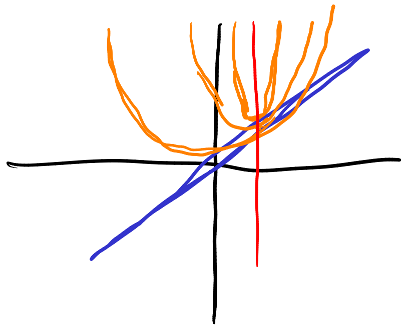
$$a\underset{\uparrow =1}{x^*} + \lambda^* = 0$$

$$\lambda^* = -a$$

Penalty method

$$\text{minimize} \quad f(x) + \rho\left(h(x)\right)^2$$

$$\text{minimize} \quad ax + \rho(x-1)^2$$



minimum at $a + 2\rho(x^{\rho*}+1) = 0$

$$x^{\rho*} = \frac{2\rho - a}{2\rho} = 1 - \frac{a}{2\rho}$$

$$h(x^{\rho*}) = x^{\rho*} - 1 = 1 - \frac{a}{2\rho} - 1 = \frac{a}{2\rho}$$

$$= \boxed{\frac{-\lambda^*}{2\rho}}$$

in general
for penalty method $\quad h \approx -\dfrac{\lambda^*}{2\rho}$

infeasibility $\downarrow \Rightarrow \rho \uparrow \Rightarrow$ poor conditioning of Hessian

Augmented Lagrangian function

$$= \sum_i (h_i(\vec{x}))^2$$

$$\hat{f}(\vec{x},\vec{\lambda}) = f(\vec{x}) + \underbrace{\vec{\lambda}^T \vec{h}(\vec{x})}_{\text{Lagrange multiplier}} + \underbrace{\frac{\rho}{2} \|\vec{h}(\vec{x})\|_2^2}_{\text{penalty term}}$$

$$\nabla_{\vec{x}} \hat{f}(\vec{x},\vec{\lambda}) = \nabla_{\vec{x}} f(\vec{x}) + \underbrace{\left(\vec{\lambda} + \rho \vec{h}(\vec{x})\right)}^T \nabla_{\vec{x}} \vec{h}(\vec{x})$$

at optimality

$$0 = \nabla_{\vec{x}} f(\vec{x}) + \lambda^* \nabla_{\vec{x}} \vec{h}(\vec{x})$$

update $\quad \vec{\lambda}^{k+1} = \vec{\lambda}^k + \rho \vec{h}(\vec{x})$

Augmented Lagrange Method

Start with

$\rho > 0$

$\gamma > 1$

$\vec{\lambda} = 0$

loop until converged

$\quad \vec{x}^* \leftarrow \operatorname{argmin} \hat{f}(\vec{x},\vec{\lambda})$

$\quad \vec{\lambda} \leftarrow \vec{\lambda} + \rho \vec{h}(\vec{x}^*)$

$\quad \rho \leftarrow \gamma \rho$

return x^*

Informally

λ converge λ^*

Thm. 17.6 in N+W

For simple example infeasibility at each step:

$$\nabla \hat{f}\Big|_{x^{a*}} = 0 = \nabla f(x^{a*}) + (\lambda + \rho h(x^{a*})) \nabla h(x^{a*})$$

$$0 = a + \lambda + \rho(x^{a*} - 1)$$

$$x^{a*} = \frac{-a-\lambda}{\rho} + 1$$

$$h(x^{a*}) = \frac{-a-\lambda}{\rho} = \boxed{\frac{\lambda^* - \lambda}{\rho}}$$

current L.M. estimate

Inequality Constraints

$$\hat{f}(\vec{x},\vec{\lambda}) = f(\vec{x}) + \vec{\lambda}^T \bar{g}(\vec{x}) + \frac{1}{2}\rho \|\bar{g}(\vec{x})\|_2^2$$

E.D.O.

$$\bar{g}_i(\vec{x}) = \begin{cases} h_i(\vec{x}) & \text{for equality constraints} \\ g_i(\vec{x}) & \text{if } g_i(\vec{x}) \geq -\frac{\lambda_i}{\rho} \\ -\frac{\lambda_i}{\rho} & \text{otherwise} \end{cases}$$

complementary slackness

(other method in N+W section 17.4)

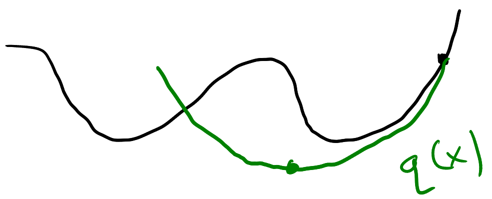
replace $g(x) \leq 0$ with $g(x) + s = 0$
$s \geq 0$

Primal-Dual Method

Basic Idea: solve KKT equations with Newton's method

Previously: Newton's method for optimization

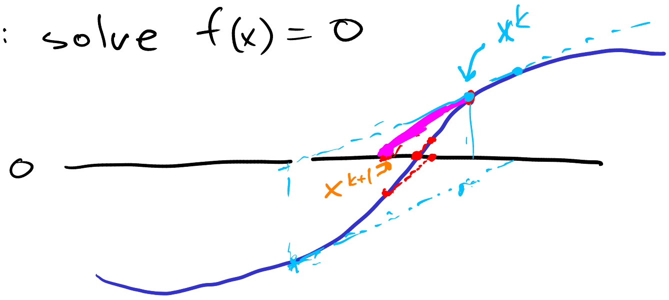
$q'(x) = 0$



$$x^{k+1} \leftarrow x^k - \frac{f'(x^k)}{f''(x^k)}$$

Now: Newton's method for nonlinear equations

Goal: solve $f(x) = 0$



Break:

$$x^{k+1} \leftarrow x^k - \frac{f(x^k)}{f'(x^k)}$$

$$f'(x^k)(x^k - x^{k+1}) = f(x^k)$$

System: $\vec{f}(\vec{x}^k)$

KKT conditions

$$\vec{x}^{k+1} = \vec{x}^k + \Delta\vec{x}$$

$$J_f(\vec{x}^k)(-\Delta\vec{x}) = \vec{f}(\vec{x}^k)$$

$$J_f(\vec{x}^k)\Delta\vec{x} = -\vec{f}(\vec{x}^k)$$

Primal-Dual Interior Point Methods

using the notation EDO

$$\begin{aligned}&\text{minimize } f(x)\\ &\text{st} \quad h(x)=0\\ &\qquad\;\; g(x)\le 0\end{aligned} \longrightarrow \begin{aligned}&\underset{x,s}{\text{minimize}}\; f(x) - \mu_b\sum_{j=1}^{n}\ln(s)\\ &\text{s.t.} \quad h(x)=0\\ &\qquad\;\; g(x)+s=0\\ &\qquad\;\; s\ge 0\end{aligned}$$

KKT conditions

Stationarity for x

L.M. for $g(x)+s=0$

$$\mathcal{L}(x,\lambda,\sigma,s) = f(x) - \mu_b \overset{\text{ones}}{e}^T\ln(s) + h(x)^T\lambda + (g(x)+s)^T\sigma$$

$$\boxed{\nabla_x\mathcal{L} = \nabla f(x) + J_g(x)^T\sigma + J_h(x)^T\lambda}$$

Stationarity for s

$$\boxed{\nabla_s\mathcal{L} = -\mu_b S^{-1}e + \sigma}$$

$$S=\begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_n\end{bmatrix} \qquad S^{-1}=\begin{bmatrix} 1/s_1 & & \\ & \ddots & \\ & & 1/s_n\end{bmatrix}$$

Feasibility

$$\boxed{\begin{aligned}h(x)&=0\\ g(x)+s&=0\end{aligned}}$$

$$\vec{F}(\vec{x}) = \begin{bmatrix}\nabla_x\mathcal{L}(x,\lambda,\sigma)\\ h(x)\\ g(x+s)\\ \sigma-\mu S^{-1}e\end{bmatrix} = 0$$

$$\Sigma=\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m\end{bmatrix}$$

$$\begin{bmatrix} H_{\mathcal{L}}(x) & J_h(x)^T & J_g(x)^T & 0\\ J_h(x)^T & 0 & 0 & 0\\ J_g(x) & 0 & 0 & I\\ 0 & 0 & S & \Sigma\end{bmatrix}\begin{bmatrix}\overset{\Delta x}{p_x}\\ p_\lambda\\ p_\sigma\\ p_s\end{bmatrix} = -\begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}$$

$$x^{k+1} \leftarrow x^k + \alpha p_x$$
$$s^{k+1} \leftarrow s^k + \alpha p_s$$
$$\lambda^{k+1} \leftarrow \lambda^k + \alpha p_\lambda$$
$$\sigma^{k+1} \leftarrow \sigma^k + \alpha p_s$$

Algorithm 5.7 Interior-point method with a quasi-Newton approximation

Inputs:

x_0: Starting point
τ_{opt}: Optimality tolerance
τ_{feas}: Feasibility tolerance

Outputs:

x^*: Optimal point
$f(x^*)$: Optimal function value

$\lambda_0 = 0$; $\sigma_0 = 0$ — Initial Lagrange multipliers
$s_0 = 1$ — Initial slack variables
$\tilde{H}_{\mathcal{L}_0} = I$ — Initialize Hessian of Lagrangian approximation to identity matrix
$k = 0$
while $\|\nabla_x \mathcal{L}\|_\infty > \tau_{\text{opt}}$ or $\|h\|_\infty > \tau_{\text{feas}}$ **do**
  Evaluate $J_h, J_g, \nabla_x \mathcal{L}$
  Solve the KKT system (Eq. 5.100) for p

$$\begin{bmatrix} \tilde{H}_{\mathcal{L}_k} & J_h^\intercal & J_g^\intercal & 0 \\ J_h(x) & 0 & 0 & 0 \\ J_g(x) & 0 & 0 & I \\ 0 & 0 & I & S^{-1}\Sigma \end{bmatrix} \begin{bmatrix} p_x \\ p_\lambda \\ p_\sigma \\ p_s \end{bmatrix} = - \begin{bmatrix} \nabla_x \mathcal{L}(x, \lambda, \sigma) \\ h(x) \\ g(x) + s \\ \sigma - \mu S^{-1} e \end{bmatrix}$$

  $\alpha_{\max} = \texttt{alphamax}(s, p_s)$ — Use Alg. 5.6
  $\alpha_k = \texttt{backtrack}(p_x, p_s, \alpha_{\max})$ — Line search (Alg. 4.2) with merit function (Eq. 5.101)
  $x_{k+1} = x_k + \alpha_k p_x$ — Update design variables
  $s_{k+1} = s_k + \alpha_k p_s$ — Update slack variables
  $\alpha_\sigma = \texttt{alphamax}(\sigma, p_\sigma)$
  $\lambda_{k+1} = \lambda_k + \alpha_\sigma s_\lambda$ — Update equality Lagrange multipliers
  $\sigma_{k+1} = \sigma_k + \alpha_\sigma s_\sigma$ — Update inequality Lagrange multipliers
  Update $\tilde{H}_{\mathcal{L}_{k+1}}$ — Compute quasi-Newton approximation using Eq. 5.91
  $\mu_b = \rho \mu_b$ — Reduce barrier parameter
  $k = k + 1$
end while