A Study on the Convergency Property of the Auxiliary Problem Principle

Balho H. Kim*

Abstract - This paper presents the convergency property of the Auxiliary Problem Principle when it is applied to large-scale Optimal Power Flow problems with Distributed or Parallel computation features. The key features and factors affecting the convergence ratio and solution stability of APP are also analyzed.

Keywords: Auxiliary problem principle, control parameter, optimal power flow

1. Introduction

There have been many mathematical decomposition coordination methods amenable to solving large-scale problems with separable structure. One of them is the so-called Auxiliary Problem Principle (APP), which was first introduced and has been extended by Cohen et al. [1, 2, 3] as a practical decomposition coordination method applicable for large-scale non-smooth, non-convex engineering problems like optimal power flow, chemical processing and transportation problems. Though its mathematical advantage and engineering applicability have been highly recognized through historical researches and expansive applications, the convergency property of the APP still remains as a subject deserving academic interests and challenge. In this paper, we present an approach on the convergency property of the APP and its variants for engineering applications.

Consider a typical convex program with separable structure of the form:

$$\text{(P)} \qquad \min_{x,z}\{f_a(x)+f_b(z): Ax=z\}. \qquad (1)$$

Then the augmented Lagrangian for problem (P) is defined as

$$\mathcal{L}\ (x,z,\lambda)=f_a(x)+f_b(z)+\lambda^+(Ax-z)+\frac{\gamma}{2}\|Ax-z\|^2, \qquad (2)$$

where λ denotes a Lagrange multiplier and γ is a constant. Augmented Lagrangians have several advantages compared to standard Lagrangians. However, the principal disadvantage for decomposition methods is the presence of the term $\frac{\gamma}{2}\|Ax-z\|^2$ in the $\mathcal{L}$, which destroys the separability between x and z, since they are linked by the cross product term z^+Ax. This has long been recognized as one of the major drawbacks of the augmented Lagrangian approach, and a number of strategies have been proposed to remove this difficulty [4-7].

In 1958, Uzawa [8] suggested to simply minimize the Lagrangian function $\mathcal{L}$ with respect to x and z (with λ fixed), then update the multiplier λ. In the method, both f_a and f_b are assumed to be strongly convex (see definition 3), and this restricts its potential applications in many interesting problems.

Gabay and Mercier [9], Tseng [7], Eckstein et al. [10, 11] proposed the alternating direction method. The basic idea underlying this approach is to sequentially perform the minimization with respect to x with z, λ fixed, then with respect to z, followed by an update of the multiplier λ. This approach removes the difficulty of the joint minimization in x and z, and thus preserves separability. This approach can be viewed as a variant of sequential decomposition techniques.

In [1, 2, 3], Cohen et al. developed a unified framework via APP which allows them to put both two-level algorithms [12, 13] and conventional optimization algorithms, e.g. gradient, Newton-Raphson, in the same context. Cohen et al. demonstrated the power and versatility of APP in the analysis and development of new decomposition algorithms. We first provide a short summary of relevant mathematical tools in the following section, followed by the key features of APP and its convergency property.

2. Preliminaries

This section summarizes some basic mathematical definitions and results that will be used in this study. All

* Dept. of Electrical Engineering, Hongik University, Seoul, Korea. (bhkim@wow.hongik.ac.kr)
Received January 2, 2006 ; Accepted September 12, 2006

these can be found in one of [14, 15]. For notational consistency, all the notations follow [15]. In addition, all the functions introduced in this section are assumed to be from a convex set C into [-∞, +∞].

Definition 1 The effective domain of a convex function $f : C \to R \cup \{-\infty, +\infty\}$ is the set

$$domf = \{x | f(x) < +\infty\}.$$

A convex function f is said to be proper if $domf$ is nonempty and

$$f(x) > -\infty, \ \forall x \in domf.$$

Definition 2 A real-valued function $f : C \to R$ is said to be lower semi-continuous at a point $x \in S \subset R^n$ if

$$f(x) = \lim_{y \to x} \ \inf f(y).$$

where $\inf f(y)$ denotes the infimum of the function f over S. (See [15], Section 7.)

Throughout the paper, f is assumed to be a convex, proper, and lower semi-continuous function.

$H \subset R^n$ denotes a space and H^f a closed convex subset of H. All other notations follow the conventional notations as in [15] or are otherwise specified.

Proposition 1 Let f be a continuously differentiable function on a convex set C with gradient ∇. Then f is convex if and only if

$$f(x) - f(y) \geq \nabla f(y)^+ (x - y), \forall x, y \in C.$$

Definition 3 If f is strongly convex, then

$$f(x) - f(y) \geq \nabla f(y)^+ (x - y) + \frac{b}{2}\|x - y\|^2$$

for some $b > 0$. (See [15], Section 25.)

Definition 4 A vector y^* is said to be a subgradient of a convex function f at a point y if

$$f(x) \geq f(y) + y^{*+}(x - y), \ \forall x \in domf.$$

The set of all subgradients of f at y is called the subdifferential of f at y and is denoted by $\partial f(y)$. Similarly, for a concave function g, the subdifferential of g is the set

$$\partial g(y) = \{y^* \mid -y^* \in \partial(-g(y))\}.$$

Proposition 2 If the subgradient of a function f at a point y is unique, then f is differentiable at y and the subgradient equals the gradient of f at y. (See [15], Section 25.)

Definition 5 An operator $Q : R^n \to R^n$ is said to be nonexpansive if and only if

$$\|Q(x) - Q(y)\| \leq \|x - y\|, \ \forall x, y \in R^n$$

and strongly nonexpansive if and only if there exists $\Theta > 0$ such that $\forall x, y \in R^n$

$$\|Q(x) - Q(y)\|^2 \leq \|x - y\|^2 - \theta\|(I - Q)(x) - (I - Q)(y)\|^2,$$

where I denotes the identity mapping.

Definition 6 The graph of an operator $T : R^n \to 2^{R^n}$ is the subset of $R^n \times R^n$

$$G(T) = \{(x, u); u \in T(x)\}.$$

Definition 7 The operator T is monotone if $(u - v)^+(x - y) \geq 0, \forall x, y, u, v \in R^n$ such that $u \in T(x)$, $v \in T(y)$. If the graph of T is not properly included in the graph of a distinct monotone operator, then T is called maximal monotone.

Proposition 3 If $T = \partial f$ is the subdifferential of a proper lower semi-continuous convex function f, then T is maximal monotone. (See [15], Section 31.)

Definition 8 The inverse of an operator is the operator defined by

$$y \in T(x) \Leftrightarrow x \in T^{-1}(y).$$

Definition 9 Let T be an operator, and $\beta > 0$. The resolvent P_β associated with T is defined by

$$P_\beta := (I + \frac{1}{\beta}T)^{-1}.$$

Proposition 4 (a) The operator P_β is a single-valued function. (b) T is maximal monotone if and only if P_β is strongly nonexpansive with $\theta = 1$. That is $\forall x, y \in R^n$, $\|P_\beta(x) - P_\beta(y)\|^2 \leq \|x - y\|^2 \|(I - P_\beta)(x) - (I - P_\beta)(y)\|^2$. Moreover the set of fixed points of P_β is equal to the set $T^{-1}(0)$ [16].

3. Analysis of Renewable Portfolio Standards

3.1 General concept of Auxiliary Problem Principle

This section presents the Auxiliary Problem Principle

(APP). We first introduce the basic idea of APP, then describe how the APP can be incorporated into the augmented Lagrangian in deriving algorithms that are suitable for parallel implementations.

Consider an optimization problem of the following form

$$\min_{u \in H^f} \{ f(u) : \theta(u) = 0 \} \tag{3}$$

where f is a convex, proper, lower semi-continuous function, with f and $\Theta(u)$ additive. Then solving the problem (3) is equivalent to solving the following sequence of auxiliary problems [1, 2]:

Algorithm - APP [2]

$$u^{k+1} = \arg\min_{u \in H^f} \{ \beta K(u) + f(u) - \beta \nabla K(u^k)^+ u + (\lambda^k + \gamma \Theta(u^k))^+ \Theta(u) \} , \tag{4}$$

$$\lambda^{k+1} = \lambda^k + \alpha \Theta(u^{k+1}) , \tag{5}$$

where K is a differentiable function.

In case that

$$u = \begin{pmatrix} u_1 \\ u_N \end{pmatrix} ; \; u_i \subset H_i \; ; \; H = H_1 \times \ldots \times H_N$$

$$; \; H_f = H_1^f \times \ldots \times H_N^f ,$$

$$f(u) = \sum_{i=1}^{N} f_i(u_i) \; ; \; \Theta(u) = \sum_{i=1}^{N} \Theta_i(u_i) ,$$

so that both f and Θ are additive with respect to a decomposition of H, then taking the auxiliary functional $K(u) = |u|^2 / 2$ yields the following subproblems, for $i = 1, \ldots, N$,

$$\min_{u_i \in H_i^f} \left\{ f_i(u_i) + \frac{\beta}{2} \| u_i - u_i^k \|^2 + (\lambda^k)^+ \Theta_i(u_i) + \gamma (\sum_j^N \Theta_j(u_j^k))^+ \Theta_i(u_i) \right\} . \tag{6}$$

For instance, setting

$$N = 2 \; ; \; u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} x \\ z \end{pmatrix} \; ; \; f = f_a + f_b$$

$$; \; \Theta(u) = Ax - z$$

identifies the problem (1). The last term in Equation (6) is the linearization of the augmented term $\frac{\gamma}{2} \|\Theta(u)\|^2$. The APP can be interpreted as solving a sequence of problems involving the linearization of the augmented Lagrangian

$$L_\gamma(u, \lambda) \equiv f(u) + \lambda^+ \Theta(u) + \frac{\gamma}{2} \|\Theta(u)\|^2 .$$

3.2 Convergence Property of Algorithm-APP

Define

$$q^k = \arg\max_{q \in H^*} : \langle q, y^k \rangle - \frac{1}{2\gamma} \| \lambda^k - q \|^2 \tag{7}$$

$$q^{k+1} = \arg\max_{q \in H^*} : \langle q, y^{k+1} \rangle - \frac{1}{2\gamma} \| \lambda^{k+1} - q \|^2 \tag{8}$$

where H^* is the dual of a Hilbert space H as noted in Section 2. Then one obtains

$$q^k = \max\{0, \lambda^k + \gamma \Theta(y^k)\} \tag{9}$$

$$q^{k+1} = \max\{0, \lambda^k + \gamma \Theta(y^{k+1})\} \tag{10}$$

and from (5)

$$\lambda^{k+1} = \lambda^k + \frac{\alpha}{\gamma} \left(q^{k+1} - \lambda^k \right) \tag{11}$$

where $\Theta(y)$ denotes the estimation of the coupling constraint at y. Back to the original problem (4), a unique solution y^{k+1} to (4) is characterized by the following variational inequality,

$$\forall y^{k+1} \in H^f,$$

$$\left\langle K'(y^{k+1} - K'(y^k), y^* - y^{k+1} \right\rangle + \frac{1}{\beta} \left(J(y^*) - J(y^{k+1}) \right) + \frac{1}{\beta} \left\langle \lambda^k, \Theta(y^*) - \Theta(y^{k+1}) \right\rangle \geq 0 \tag{12}$$

Then by (9),

$$\left\langle \lambda - q^k, \Theta(y^k) + \frac{1}{\gamma} (\lambda^k - q^k) \right\rangle \leq 0, \forall \lambda \in H^* \tag{13}$$

and, furthermore, rewriting (12) for iteration (k+1), we obtain

$$\left\langle \lambda - q^{k+1}, \Theta(y^{k+1}) + \frac{1}{\gamma} (\lambda^{k+1} - q^{k+1}) \right\rangle \leq 0, \forall \lambda \in C^* . \tag{14}$$

In particular, for $\lambda = \lambda^*$ we obtain

$$\left\langle \lambda^* - q^{k+1}, \Theta(y^{k+1}) + \frac{1}{\gamma}(\lambda^{k+1} - q^{k+1}) \right\rangle \le 0, \forall \lambda \in H^*. \quad (15)$$

Using (11), one gets

$$\left\|\lambda^{K+1} - \lambda^*\right\|^2 = \left\|\lambda^K - \lambda^*\right\|^2 + \frac{\alpha^2}{\gamma^2}\left\|q^{K+1} - \lambda^k\right\|^2 + \frac{2\alpha}{\gamma}\left\langle \lambda^k - \lambda^*, q^{k+1} - \lambda^k \right\rangle. \quad (16)$$

Next, add (2α)*(13) to (15) to get

$$\left\|\lambda^{K+1} - \lambda^*\right\|^2 \le \left\|\lambda^K - \lambda^*\right\|^2 + \frac{\alpha(\alpha - 2\gamma)}{\gamma^2}\left\|q^{K+1} - \lambda^k\right\|^2 + 2\alpha\left\langle \Theta(y^{k+1}), q^{k+1} - \lambda^k \right\rangle. \quad (17)$$

Since

$$\left\langle q^k, \Theta(y^*) \right\rangle \le \left\langle \lambda^*, \Theta(y^*) \right\rangle, \quad (18)$$

one obtains the following inequality by adding (17) to (18) and rearranging terms

$$\left\|\lambda^{K+1} - \lambda^*\right\|^2 \le \left\|\lambda^K - \lambda^*\right\|^2 + \frac{\alpha(\alpha - 2\gamma)}{\gamma^2}\left\|q^{K+1} - \lambda^k\right\|^2 + 2\alpha\left\langle \Theta(y^{k+1}), q^{k+1} - \lambda^k \right\rangle + 2\alpha\left\langle \Theta(y^{k+1}) - \Theta(y^*), q^k - \lambda^* \right\rangle, \quad (19)$$

where, in (19), we made use of

$$L(y^*, \lambda) \le L(y^*, \lambda^*) \le L(y, \lambda^*), \ \forall y \in H^f, \ \forall \lambda \in H^* \quad (20)$$

with setting $\lambda = q^k$. Then the variational inequality given in (12) can be rewritten as

$$K(y^{k+1}) - K(y^k) + \left\langle K'(y^{k+1}), y^* - y^{k+1} \right\rangle - \left\langle K'(y^k), y^* - y^k \right\rangle + \frac{1}{\beta}\left[J(y^*) - J(y^{k+1})\right] - \frac{b}{2}\left\|y^{k+1} - y^k\right\|^2 + \frac{1}{\beta}\left\langle \lambda^k, \Theta(y^*) - \Theta(y^{k+1}) \right\rangle \ge 0, \quad (21)$$

where we made use of the strong convexity of $K(y)$ with constant b.

Let us define a function $\psi(y, \lambda)$:

$$\psi(y, \lambda) = K(y^*) - K(y) - \left\langle K'(y), y^* - y \right\rangle + \frac{1}{2\eta}\left\|\lambda - \lambda^*\right\|^2 \quad (22)$$

where $\eta > 0$. Then again by the strong convexity of $K(y)$,

$$\psi(y, \lambda) \ge 0.$$

For convenience, define

$$\phi^k = \psi(y^k, \lambda^k)$$

and

$$\phi^{k+1} = \psi(y^{k+1}, \lambda^{k+1})$$

Then

$$\phi^{k+1} - \phi^k = -K(y^{k+1}) + K(y^k) - \left\langle K'(y^{k+1}), y^* - y^{k+1} \right\rangle + \left\langle K'(y^k), y^* - y^k \right\rangle + \frac{1}{2\eta}\left\{\left\|\lambda^{k+1} - \lambda^*\right\|^2 - \left\|\lambda^k - \lambda^*\right\|^2\right\} \quad (23)$$

From RHS of (20) with $y = y^{k+1}$, we get

$$J(y^*) + \left\langle \lambda^*, \Theta(y^*) \right\rangle \le J(y^{k+1}) + \left\langle \lambda^*, \Theta(y^{k+1}) \right\rangle$$

or,

$$\frac{1}{\beta}\left[J(y^*) - J(y^{k+1})\right] \le \frac{1}{\beta}\left\langle \lambda^*, \Theta(y^{k+1}) - \Theta(y^*) \right\rangle.$$

Next add $(\frac{1}{2\alpha\beta})$*(19) to (21) to get

$$\phi^{k+1} - \phi^k \le -\frac{b}{2}\left\|y^{k+1} - y^k\right\|^2 + \frac{1}{\beta}\left\langle q^{k+1} - q^k, \Theta(y^{k+1}) \right\rangle + \frac{(\alpha - 2\gamma)}{2\alpha\gamma^2}\left\|q^{k+1} - \lambda^k\right\|^2 \quad (24)$$

We define another function $\gamma(y, \lambda)$ as:

$$\gamma(y,\lambda) = \psi(y,\lambda) - \frac{\gamma}{2\beta}\|\Theta(y)\|^2 \quad (25)$$

Then from the assumption on the equality constraint y being Lipschitz with constant τ and (13) with $q^{k+1} \to \lambda$, we finally get the following relationship,

$$\gamma(y^{k+1},\lambda^k) - \gamma(y^k,\lambda^k) \le \frac{1}{2\beta}(\gamma\tau^2 - b\beta)\|y^{k+1} - y^k\|^2 + \frac{1}{2}\beta(\alpha - 2\gamma)\|\Theta(y^{k+1})\|^2 . \quad (26)$$

Thus with the following assumptions

(a) $\frac{1}{2\beta}(\gamma\tau^2 - b\beta) < 0$ and

(b) $(\alpha - 2\gamma) < 0$,

one can verify that the function $\gamma(y,\lambda)$ is non-increasing.

Next, combining the following

$$\|\Theta(y)\| = \|\Theta(y) - \Theta(y^*)\| \le \tau\|y - y^*\|$$

with the assumption that the functional K be strongly convex with the constant b, obtains

$$\gamma(y,\lambda) \ge \frac{(b - \varepsilon c\tau^2)}{2}\|y - y^*\|^2 + \frac{\varepsilon}{2}\|\lambda - \lambda^*\|^2 .$$

That is the function $\gamma(y,\lambda)$ bounded below.

Consequently,

$$\|y^{k+1} - y^k\|^2 \to 0$$

and

$$\|\Theta(y^{k+1})\|^2 = \|y_a^{k+1} - y_b^{k+1}\|^2 \to 0 .$$

This proves the convergency of APP. (Ref. Fig. 1.) Finally, with the identification of $b = 1$, and $\tau = 1$, we obtain the following convergence condition,

$$\beta > \gamma > \frac{\alpha}{2}$$

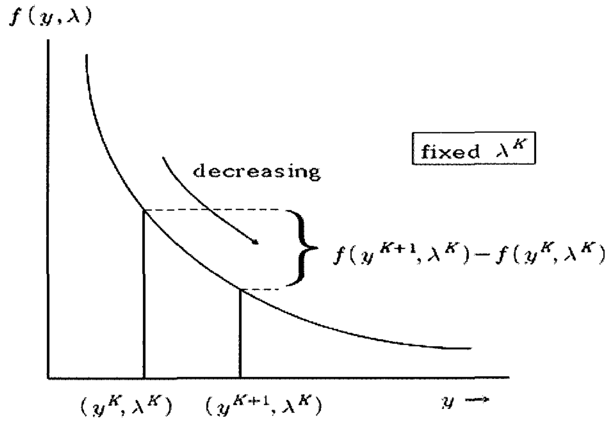

Fig. 1. Convergency Property of APP

3.3 Choice of Parameters

In Algorithm-APP, as also commented in [2], our experience shows that the choice of $\alpha = \gamma$ seems the best among other possible combinations. However, in choosing the parameter β, one must consider the relationship with α and γ. The convergence conditions given in the previous section can be rewritten as

(a) $\frac{\gamma}{\beta} - 1 < 0$,

(b) $\frac{(\alpha - 2\gamma)}{2\beta} < 0$,

and to enhance the convergence ratio, one might choose α, β, γ so that the negative values of LHS of (a) and (b) are as large as possible. Our experience shows that the value of LHS of (a) dominates the convergence.

In our study, we mostly adopt the following relationship,

$$\alpha = \gamma = \frac{\beta}{2},$$

where the average of 2nd order coefficient over all the core generators in the two adjacent regions is taken as the value of β. Our studies found that the choice of parameters depends on the problems. The applications and further studies on the choice of parameters can be found in [19, 20, 21].

4. Conclusion

This paper presents the convergency property of the Auxiliary Problem Principle (APP) which has already been applied to various engineering optimization problems with separable structure such as Distributed Optimal Power

Flow and Parallel Economic Dispatch, etc. The significance of this study lies in the fact that the mathematical property of the convergency has been proved in a different way based on the existing decomposition methodologies. The choice of parameters for improving convergence ratio in practical application remains for future study.

Acknowledgements

This work was funded by Ministry of Education and Human Resources Development as for BK21 (2nd Phase) Project 'Research on New Energy Resource & Power System Interface' and done during the sabbatical year (09/01/2005 – 08/31/2006).

References

[1] Guy Cohen, "Optimization by decomposition and coordination: A unified approach," *IEEE Transactions on Automatic Control*, AC-23(2), pp. 222-232, April 1978.

[2] Guy Cohen, "Auxiliary problem principle and decomposition of optimization problems," *Journal of Optimization Theory and Applications*, 32(3), pp. 277-305, November 1980.

[3] Guy Cohen and Bernadette Miara, "Optimization with an auxiliary constraint and decomposition," *SIAM Journal on Control and Optimization*, 28(1), pp. 137-157, January 1990.

[4] D. P. Bertsekas, "Convexification procedures and decomposition methods for nonconvex optimization problems," *SIAM Journal on Control and Optimization*, 49(2), pp. 169-197, October 1979.

[5] E. Gelman and J. Mandel, "On multilevel iterative methods for optimization problems," *Mathematical Programming*, 48, pp. 1-17, 1990.

[6] E. V. Tamminen, "Sufficient conditions for the existence of multipliers and lagrangian duality in abstract optimization problems," *Journal of Optimization Theory and Applications*, 82(1), pp. 93-104, July 1994.

[7] P. Tseng, "Applications of a splitting algorithm to decomposition in convex programming and variational inequalities," *SIAM Journal on Control and Optimization*, 29(1), pp. 119-138, January 1991.

[8] H. Uzawa, *Studies in Linear and Nonlinear Programming*, Springer-Verlag, 1958.

[9] D. Garby and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite-element approximation," *Comp. Math. Appl.*, 2, pp. 17-40, 1976.

[10] J. Eckstein, "Parallel alternating direction multiplier decomposition of convex programs," *Journal of Optimization Theory and Applications*, 80(1), pp. 39-63, January 1994.

[11] J. Eckstein and P. D. Bertsekas, "On the douglas-rachford splitting method and the priximal point algorithm for maximal monotone operators," *Mathematical Programming*, 55(3), pp. 293-318, 1992.

[12] M. D. Mesarovic, D. Macko, and Y Takahara, *Theory of Hierarchical, Multilevel, Systems,* Mathematics in science and engineering, Academic Press, New York, 1970.

[13] Y. Takahara, *Multilevel Approach to Dynamic Optimization*, SRC-50-C-64-18, Case Western Reserve University, Cleveland, Ohio, 1964.

[14] Leon S. Lasdon, *Optimization Theory for Large Systems*. The Macmillan Company, New York, 1970.

[15] R. T. Rockefeller, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.

[16] R. T. Rockefeller, "Monotone operators and the proximal point algorithm," *SIAM Journal on Control and Optimization*, 14(5), pp. 877-898, August 1976.

[17] C. B. Brosilow, L. S. Lasdon, and J. D. Pearson, "Feasible optimization methods in inter-connected systems," *Prec, JACC*, Troy, NY, 1965.

[18] G. B. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Operations Research*, 8, January 1960.

[19] D. Hur, J. K. Park, and B. H. Kim, "Evaluation of convergence rate in auxiliary problem principle for distributed optimal power flow," *IEE Proceedings-Generation, Transmission & Distribution*, 149(5), pp. 525-532, September 2002.

[20] B. H. Kim and R. Baldick, "A comparison of distributed optimal power flow algorithms," *IEEE Transactions on Power Systems*, 15(2), pp. 599-604, May 2000.

[21] D. Hur, J. K. Park, and B. H. Kim, "On the convergence rate improvement of mathematical decomposition technique on distributed optimal power flow," *International Journal of Electrical Power & Energy Systems*, 25, pp. 31-39. January 2003.

Balho H. Kim
He received his B.S. degree from Seoul National University, Korea, in 1984, and his M.S. and Ph.D. degrees from the University of Texas, Austin in 1992 and 1996, respectively. He was with KEPCO (the Korea Electric Power Corporation) from 1984 to 1990 and then joined Hongik University in 1997 where he is presently an Associate Professor of Electrical Engineering.